UNIVERSITY OF LIVERPOOL

Selective Sharing; A dialogical approach to Ontology Matching

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Gabrielle Santos

February 2018

Contents

List of Figures

List of Tables

Preface

This thesis is primarily my own work. The sources of other materials are identifed.

Abstract

Ontologies are used by agents to model knowledge for their representation of a world. In order for different agents to communicate with each other, it is necessary to align their respective ontologies. Due to the heterogeneous nature of these representations, this requires that a common meaning is established to allow accurate and successful communication over a given task.

Traditional ontology alignment techniques rely on the full disclosure of ontological models, which aim to find the optimal set of correspondences to map entities from one ontology to another. Due to privacy concerns amongst other issues, these approaches may not always be pragmatic or acceptable in open and opportunistic environments. Despite this, within the ontology matching community, development of alignment techniques in order to align a full ontology, has focused on centralised approaches. Limited attention has been given to addressing these issues through application of a decentralised approach, capable of avoiding full disclosure of data.

In this thesis a novel dialogue is presented, which allows for selective data sharing during ontology alignment between agents, via a decentralised negotiation mechanism. This negotiation mechanism enables strategic agreement over correspondences between agents with limited or no prior knowledge of their opponent's ontology. Therefore both agents are able to reach a mutual agreement over an alignment through selective disclosure of their ontological knowledge and their specific strategies. In this thesis, the dialogue mechanism is formally introduced, with discussion of its behaviour, properties and outcomes.

Acknowledgements

"You have to stay faithful to what you're working on."

—Stephen King

This PhD has been an incredible opportunity allowing me to evolve my skills and interests, canvassing a multi-interdisciplinary background to develop and painstakingly navigate the world of research. The resulting thesis has been a tremendous labour of love and at times, something far less than love. It has however, been a pleasure and a privilege to complete and something I am immensely proud of. There are many people without whom I could not have achieved any of the work that has been presented here, so I would like to take the time to thank them for their limitless help.

I would like to gratefully acknowledge the support from my supervisors Valentina Tamma and Terry Payne, for all their infinite patience and guidance through the development of this work. Your wisdom and challenges through endless engaging discussions have provided me the opportunity to establish skills and knowledge without which this work could never have developed. *Grazie per tutto il tuo aiuto e supporto!*

To those friends from in and out of university, who nodded and smiled as I discussed this thesis and for generally being top pals. To my office mates, who could never be trumped for their support and humour during the moments of lapsing continuation of productivity. To David Geleta and Richard Williams, who with their expertise and indispensable time, drew me out of the dark hollows of research which a PhD student knows only too well.

Most importantly I would like to thank my Mum and Dad, to which this thesis is lovingly dedicated. Your ongoing support and limitless belief in me has been invaluable throughout my lifetime, providing me with infinite opportunities and love, without which I would not be the person I am today, and for this I am tremendously grateful.

Honourable mentions: Costa, Pardon Us, Wikipedia, Kanye, 92 Degrees Coffee, Cinema Sundays, Andrew W.K, Grandma, Starbucks, Member Berries, Tim Berners-Lee and Science.

"We have to remember what's important in life, friends, waffles, and work.
Or waffles, friends, and work. But work comes third"

—Leslie Knope

Part I

Introduction

Chapter 1

Background and Context

Chapter Outline

'There are some things one can only achieve by a deliberate leap in the opposite direction'. - Franz Kafka

The aim of this thesis is to investigate if an alignment can be made between two ontologies, using a dialogue based approach, where the knowledge is not shared by the agents a priori. *This chapter sets out to outline the background and context of the research presented in this thesis. This is followed by the motivation behind the research and a detailed overview of the structure of the thesis presented as a guide for the reader.*

1.1 Background and Motivation

In recent years, the volume of data is vastly increasing, largely due to the developments of the internet and social media services [22, 55], which are generating large amounts of data, with ongoing analytical opportunities. [121]

This data is being created and collected at a record rate [48, 128], often by independently run services. This significant increase in the amount of data and its diversity has given rise to the importance of understanding this data, using services and systems currently available over the Internet [15].

These independent services usually commit to their own knowledge model (*ontologies*) and interoperate in an opportunistic fashion in order to perform a given task. However, as these data models differ, the extent to which the messages are understood can be restricted; thus approaches are necessary to support semantic reconciliation and thus enable seamless interactions to take place between these services. Usually these approaches depend on reaching some form of agreement on the choice of mappings or *correspondences* to translate between the entities in two ontologies.

Ontologies are machine readable specifications of a conceptualisation of a given domain knowledge [58]; they define the entities and the relationships between them modelling a domain. In review of the definitions presented in this work, ontologies are used as in the form of RDF graphs assigned to an agent and are defined as a formal, explicit specification of a partially shared conceptualisation. The ability to reconcile independently developed knowledge sources is crucial in supporting critical decision making in intelligent applications that require the interaction between disparate knowledge sources.

Due to the subjective nature of ontological design, there is always multiple interpretations for the modelling of knowledge in a domain. It is therefore unlikely that two agents would model the same domain using the same terms and levels of detail. These modelling differences of the ontologies lead to heterogeneity and can be costly when agents are addressing the task at hand, and can lead to miscommunication between the systems.

With the diversity and volume of data expected to further increase in the near future, collaboration and co-ordination between different systems is essential. This has led to a pressing demand for more flexible knowledge-based intelligent applications (*agents*) that can exploit this machine-readable domain knowledge (*ontologies*), and perform tasks that require the integration of disparate knowledge sources.

Within the multi agent system area, several research efforts have investigated the theoretical underpinning of the mechanisms that support different forms of interactions (e.g. co-ordination, negotiation, etc). The success of such interaction mechanisms depends on the agents ability to successfully make sense of the content of the messages they exchange. However, the increasing number of independent knowledge-based services that have appeared has exacerbated the problems caused by the use of heterogeneous vocabularies for representing the content of communication messages.

In order for the agents to complete a task successfully, communication and accurate interaction is critical, and thus it can be cast as an autonomous AI problem. Approaching communication in terms of a dialogue game provides a protocol based on a set of moves rooted in speech acts theory [10]. The protocol provides a formalised approach of decentralised communication between two participating agents, over a goal or task to be achieved. This task of mutually understanding the knowledge models can be approached by mapping these ontologies, however this is not always a straight forward task.

Ontologies are designed for specific tasks, or for a particular domain, therefore can vary greatly in terms of structure, granularity and nature of design. This subjectivity causes problems regarding the meaning of terms when the two agents try to communicate. Take for example two individuals each with a different mother tongue. They are gathered together both to discuss a concept label they both identify as 'sport', however they both have very differing languages describing the given domain, and the structure in which they categorise the subject. This is an example of heterogeneity, and can give rise to mismatched concepts, if not properly resolved. In order for agents to successfully communicate over their committed ontologies, the knowledge bases are inputted into a mapping process. This mapping process allows the concepts in the ontologies to be mapped to a candidate concept in the other ontology in order to generate an alignment over the two.

Mapping ontologies is not always an elementary process, and it is often complicated by heterogeneity giving rise to the ontology matching problem, where agents need to communicate with each other, yet commit to different and often varyingly defined ontologies. This task is exasperated by semantic or syntactic heterogeneity in the design of ontologies, which makes the mapping process difficult, thus this complication needs to be taken into account and resolved. Ontologies are heterogeneous if they contain different characteristics and make different assumptions regarding their domain knowledge, which can be found either in the semantic or syntactic level of the ontology design.

Whilst the problem of determining the vocabulary to use when integrating heterogeneous knowledge has been investigated by numerous research efforts [13, 117, 126], they typically require that both ontological models are shared in full with some centralised party responsible for discovering the correspondences, even though there may be no guarantee that such correspondences exist.

An increase in the volume of data has developed the argument of privacy versus data use, in which there must be a balance between the analytics of data and use of the data collected, alongside the opportunities and benefits that are gained from this open collection of data, and to society itself [102, 119]. As a result of this data collection, privacy has become increasingly pertinent [20, 75] in generating ontology alignments, whereby neither agent is necessarily prepared to disclose its *full* ontology, presenting further challenges for alignment success. The notion of privacy preserving information sharing has been advocated by a number of previous efforts [1, 29, 30, 93], which include use-cases that require some form of privacy preservation in either the schema or data (or both).

This increase in the importance of privacy, is a result of the fact that it is not always necessary to align the full knowledge represented in the ontological entities, particularly in those scenarios where ontologies have grown considerably due to the inclusion of several different ontological models, or where the disclosure of certain modules in the ontologies may be problematic or undesirable, as they contain definitions that are private or commercially sensitive [56].

Various cases for which it is necessary to introduce some form of privacy preservation in the schema, in the data or both have been identified [1, 29], including monitoring healthcare crises, facilitating e-commerce, outsourcing, and end-to-end integration. More recently, the notion of preserving privacy when matching schemas and ontologies was proposed to facilitate interoperation between different parties when the possibility of sharing ontological knowledge is reduced whilst limiting the sharing of information concerning the ontologies used to model the different applications [30, 93].

Likewise, from a game-theoretic perspective, an agent may want to keep part (or all) of its ontological knowledge private, and consequently may not want to share or disclose it to other agents, as the disclosed ontological axioms could be exploited by other, self-interested agents (and thus have intrinsic value to the owner if kept private), were agents to compete over multiple transactions. For these reasons, this thesis presents a novel decentralised approach using negotiation through the use of a dialogical process. By exploring how dialogue protocols can be used to determine mappings that satisfy each of the agents requirements and strategies, this thesis presents a recasting of the ontology alignment problem. The use of dialogical models allow the agents to state their position regarding the correctness of some mapping in an asynchronous and distributed fashion, whilst maintaining control over the type of knowledge (class labels vs. ontological model) disclosed. The main contribution of this thesis is a dialogue based negotiation mechanism that allows the agents to propose viable lexical mappings and then support these proposals with evidence in the form of ontological fragments, thus collaboratively generating a mutually acceptable partial alignment. These are shared on a per-need basis, and hence the mechanism is purely opportunistic.

This dialogue process developed from principles of human conversation, and speech acts, provides an approach to facilitate agents in generating an ontology alignment.

This dialogical approach sets out rules of participation which must be adhered to by the participating agents, centred in various conversational theories including strict conversation rules defined by Pask [104], and the co-operation principle detailed by Grice. The dialogical approach also formalises the available moves the agents can make throughout the alignment process, which have been designed using the notion of speech acts developed by Austin and Searle.

This dialogical approach however, still needs to address the heterogeneity between the ontologies, it is here the novel contribution of this work is introduced. The semantic meaning of concepts is established by investigating a path in the ontology, represented as a graph, in order to resolve heterogeneity and avoid mismatches within an alignment. This unique view point has dictated how the research areas from ontologies and dialogue, detailed in the literature review of this thesis, come into play in this work, and has defined how the research has developed. The formalisms of this dialogue will be detailed in Chapter 5 with respects to the dialogue protocol, the agents strategy, and followed with an implementation of the dialogue designed to address the research question.

In contrast to current ontology matching methods using a full a priori sharing approach of the participating agent's ontologies, this work takes a more incremental approach to ontology sharing. The focus of this work is to allow the agents an element of privacy in the ontology matching problem. It is proposed in this work, that the agents do not share their full ontologies prior to participating in generating an alignment, with another agent. Here the agents share fragments of their ontology related to a given concept as part of a candidate mapping.

This novel approach to ontology matching using an incremental sharing within a dialogical approach, is detailed later in this Part, it is important here to note the way in which the ontologies are shared between the agents, when they try an develop an alignment between the concepts in their ontologies. In addressing heterogeneity within ontology alignments, this work focuses on the ability for agents to use only a fragment of the ontology in generating an alignment rather than using the ontology as a whole.

1.2 Research Aims and Contribution

This section summarises the aims and objectives of this thesis, and the contributions to the state of the art that were made in order to achieve them. The research aim of this thesis is to investigate the ability of a decentralised approach using negotiation, to generate a meaningful ontology alignment between the independent knowledge bases of two participating agents. This can be summarised by the following research question:

Can a meaningful alignment be generated over two ontologies, when knowledge is not shared a priori?

In order to investigate the above research question, this thesis set out to achieve the following goals:

i To develop a decentralised ontology matching approach, based on a dialogical protocol, allowing two participating agents to attempt to generate a meaningful alignment between their knowledge bases. Where a meaningful alignment is supported by semantic and syntactic knowledge enhancing the meaning of the correspondence found.

ii To allow the participating agents to utilise this designed protocol and generate a meaningful alignment without the agents having to share all of the knowledge within their ontology to either:

  - each other *or*;
  - to a third party centralised agent.

Using the research question detailed above and the goals outlined, the main research direction investigates the development of a dynamic approach to incremental selective sharing, using a formalised dialogical approach in order to generate these meaningful ontology alignments. Thus the novel contributions of this thesis are:

i A dialogue protocol designed specifically to allow two agents to participate in dialogical communication in order to generate an alignment over their two ontologies, addressing the ontology alignment problem. This dialogical process generates an incremental approach for determining the ontology alignment, where the agents reason collaboratively on the plausibility of establishing correspondences between entities in one agent's ontology and those in another.

ii An agent strategy designed for the agents to specify their choice of moves through the dialogue and its negotiation and approach to sharing, allowing the agents to control what information is shared dynamically throughout the dialogue, and the selection process for correspondences to be accepted into the negotiated alignment.

iii An incremental selective sharing process within this dialogue game structure, allowing the agents to garner meaning over terms without sharing their knowledge base in full.

Despite the extensive approaches designed to address the ontology alignment problem, few have approached it with the concept of privacy in mind, using a decentralised approach of ontology matching within a dialogue protocol. As this contribution is centred in the ontology community and focuses on the alignment problem, this approach presented here casts this using dialogue games, and element of negotiation, rooted in philosophy and the importance of communication between participating agents in order to successfully achieve a task.

1.3 Thesis Structure

The thesis presented here is divided into four parts and further separated into 9 chapters, concluding with an appendix. Part I introduces the background and context, detailing the motivation and the aims and contribution of this thesis, and presents the chapter structure. Part II presents the literature review and the main subjects covered in this research. Part III introduces the contribution of this thesis, based on the subject areas specified in Part II, introducing how these elements fit into the work presented followed by the main contribution. Finally Part IV presents the summary of this thesis, with a concluding chapter outlining the further work and review of the contribution. This section below provides a further breakdown of each of the chapters in this thesis and are presented as follows:

C.1 **Chapter 1** - This chapter introduces the motivation of this thesis and the background on which it has been based. Also presented here are the research aims of the work and an introduction to the contribution. Finally this chapter outlines the thesis structure including the chapters presented and the related publications based on the contributions of this thesis.

C.2 **Chapter 2** - This chapter introduces the first of the main fields covered in the literature review in addressing the notion of ontologies, beginning with their origin in philosophy to their use in computer science. In this chapter ontologies are introduced and the definitions of this concept is discussed and formally defined. This chapter then details the components of an ontology, which are integral to how ontologies are used and addressed within ontology mapping introducing the focus of the following chapter.

C.3 **Chapter 3** - This chapter follows on from the notion of ontologies defined in Chapter 2 and introduces the concept of an ontology alignment. The types of ontology alignment are then defined and discussed in terms of the relations and modelling, introducing the problem of heterogeneity. This chapter then presents the ontology alignment problem arising due to various types of heterogeneity and mismatches which are detailed. This is followed by an overview and comparison of current alignment systems.

C.4 **Chapter 4** - This chapter is the final part of the literature review and background, and presents the last of the main fields used in this research. This chapter highlights the importance of communication, in sending and receiving messages successfully and introduces dialogues and dialogue games and their uses in multi agent systems. The rules of formulating and adhering to the order of conversation are presented where the agents must abide by these rules within an interaction. The concept of speech acts are introduced and the development of moves used by agents in order to communicate. It is here, that the ontology alignment problem is cast as a dialogue game, introducing the key research question of this thesis.

C.5 **Chapter 5** - This chapter presents the main contribution of the novel dialogue protocol used to address the ontology matching problem, in the form of a dialogue game. This protocol is detailed in terms of the moves available to the agents throughout the dialogue process, including the formal definitions of these moves and the pre and post conditions, which make a move available to an agent. The mechanisms defining the interaction between the two agents, are defined in terms of utilising both lexical and structural similarity methods, in addressing the ontology alignment problem. This chapter also introduces the use of a commitment store within this dialogue protocol, and defines a proof of termination for the approach. This chapter introduces the agent decision making strategies defining the move an agent will select, at a given state in the dialogue. This chapter is finalised by detailing the variants of the protocol and strategies that are presented.

C.6 **Chapter 6** - This chapter begins with a recap of the key features of the dialogue protocol presented in Chapter 5. This is followed by a detailed and comprehensive walkthrough example of the dialogue protocol used to illustrate the dialogue process mapping concepts from two trivial ontology fragments between two participating agents. This walkthrough example is divided into two parts, firstly beginning with two agents starting a new dialogue, and the second part the same agents re-starting the dialogue over a second concept to be mapped between them. Each of these walkthrough parts are summarised with a table illustrating the moves and messages exchanged between the agents, and the knowledge shared and stored between them.

C.7 **Chapter 7** - Here the strategic decision making criteria used by the agents throughout the dialogue have been defined and assumed based on the introduction in Chapter 5. In this chapter, these metrics explicitly defined and a justification of the lexical similarity metric used within the dialogue is also presented. The metrics definitions are followed by the defined ranking system both of which are used within the implementation of the approach in Chapter 8.

C.8 **Chapter 8** - This is the final chapter in the contribution part of this thesis and presents the evaluation of the dialogue approach, beginning with the specific protocol and strategy used to evaluate the approach over real world ontologies. This chapter formally presents the detailed metrics and parameters used within the approach, and introduces the real world datasets which have been used for the evaluation. This chapter then introduces the hypotheses and research questions used, in order to develop the experiments and evaluate the dialogue approach in generating alignments. Next the three varied versions of the dialogue approach are empirically evaluated in comparison to a benchmark alignment. This evaluation is then summarised by comparing approach to current alignment systems used in the research field in addressing ontology alignment.

C.9 **Chapter 9** - This chapter is within the synopsis of this thesis and details the conclusions of the research presented. Here the contribution of this work is discussed and reviewed and possible avenues for future work to expanding this research beyond the scope of this thesis are presented.

Appendix - This thesis also includes the following appendices:

A.1 Presents the results of over all the datasets run in these experiments presented in Chapter 8. These results present the individual results of the evaluation presented for the first two variants of the DbMN approach (DbMN_5 and DbMN_6).

A.2 Presents the individual results of the all the experiments run within the evaluation of the last variant of the DbMN approach (DbMN_7).

B.1 Appendix B documents the benchmarks used in the evaluation of the DbMN approach, for each of the data set pairs. The benchmarks detail the differences between the *gold* and *platinum* standards.

C.1 Presents an alternate representation of move semantics for the dialogue protocol based on the PARMA protocol [9].

D.1 This Appendix lists a glossary of symbols assumed for the purposes of this thesis.

This thesis references work that has been previously published, and also work which is currently up for review, these are detailed as follows:

- *In submission:* Santos. G, Tamma. V, Payne. T and Grasso. F. (2017) *Dialogical Mechanisms for Generating Correspondences between Heterogeneous Agents* Submitted to: JAAMAS;

- Santos. G, Tamma. V, Payne. T and Grasso. F. (2016) *Discovering Ontological Correspondences Through Dialogue.* In: Knowledge Engineering and Knowledge Management, EKAW 2016 Vol. 10024 (pp. 543-560) [112];

- Santos. G, Tamma. V, Payne. T and Grasso. F. (2016). *A Dialogue Protocol to Support Meaning Negotiation.* In: The 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS1026) pp 1367-1368 [114];

- Santos. G, Tamma. V, Payne. T and Grasso. F. (2015). *Dialogue Based Meaning Negotiation.* In: The 15th Workshop on Computational Models of Natural Argument [113].

Part II

Literature Review

Chapter 2

Literature Review: Ontologies

Chapter Outline

'Being qua being' - Aristotle

This chapter marks the beginning of the the literature review and introduces a key component central to this work; 'ontologies'. This chapter details and defines the concept of ontologies, beginning with what they are, and their origins rooted in philosophy, modelling the notion of 'being' *and defining what truly exists. This philosophical definition however, is not the focus of the use of ontologies in this work. It is here the transition from the philosophical uses of ontologies is developed into artificial intelligence, The uses of ontologies within computer science detailed and explored, alongside the principles of ontologies and their components which represent knowledge of a conceptualisation of a given world.*

2.1 Ontologies

The use of the term *ontology* has developed since its inception in philosophy, and has found its way into artificial intelligence with the expansion of linked data and the development and growth of the semantic web. It is important that the meaning of ontologies is examined, and a sound understanding of what they are is developed to define their components, and the context in which they are utilised in this work. Ontologies represent the vocabulary for a specific domain and are specified explicitly using concepts and relations, providing the constraints to define an entity which is being modelled. At a high level the term ontology is an *'inter-related vocabulary'* [125] of terms with a given specification of their meaning.

The term ontology takes its origin from the greek *'onto'*, where its roots are centred in philosophy, specifically, in the study of 'being'. Aristotle presented a systematic account of the study of 'being' in *metaphysics* by classifying anything that exists, in relation to what can be asserted or predicated.

In the philosophical context, entities and their relations are examined, with regards to their respective subject hierarchies, where they can be categorised and rationally understood [107]. The use of the term ontology differs across multiple disciplines: Guarino discusses these differences by distinguishing between the use of the term in philosophy and in computer science. Guarino states that the philosophical use of the term 'Ontology' is an uncountable noun which is used as the 'nature and structure of things', independently of further considerations or if what they describe even exists. In contrast to this, Guarino refers to the term 'ontology' as a countable noun, within the computer science discipline as formal models of a structure of a system, represented with relevant entities and relations pertinent to a given purpose [60].

Quine, in relation to the study of 'being', developed the 'ontological question' of *'what is there?'* [127] and used ontology as the study of answering this question. Although the central concept in Quine's use of ontology is the notion of 'being', there are similarities in his use of an ontological question, and his use of an ontology in answering this question mirrors modern ontological design principles. Competency questions such as 'what is there?' provide a way to determine the boundaries and the entities to model. The philosophical area of ontology is still the subject of research efforts, however within artificial intelligence, the meaning takes a difference stance. The focus is not in metaphysics and the definition of 'being' or the meaning of existence, but rather what is required by a system to reason in order to complete a given task.

The field of computing was quick to understand the importance of philosophy, for example Turing, and his Imitation Game [123], routed in the philosophical work of Descartes [31], famously looking at the notion of intelligence and the idea of whether machines could think independently.

McCarthy, in coining the phrase 'Artificial Intelligence' *(AI)* [88], stated clearly that some elements of philosophy in terms of human intelligence, would be crucial in the field of computing. McCarthy highlights the importance of philosophy and human behaviour in computing [89] and states that:

'To ascribe certain beliefs, knowledge, free will, intentions, consciousness, abilities or wants to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person'.

McCarthy emphasises the need for a system to understand the world in which it is operating, in terms of concepts and relations, in order for it to complete a given task.

As technology advances, more is continually expected from *AI* systems and the boundaries of intelligence are being pushed, regarding the meaning of intelligence and also what is required by a system, in order for it to act and infer as a human. Systems require an understanding of the environment they are utilising, in order to carry out more complex tasks. This emphasises the need for an unambiguous definition of an environment, so that all participants can be clear on the terms of the environment and the objects within it. Ontology concepts need to be explicit and unambiguous in their identification introducing a well cited early definition by Gruber [58] of an ontology, reflecting the aspects of ontology as a philosophical artefact:

'An ontology is an explicit specification of a conceptualization'

This definition introduces the importance of clarity and preciseness such that ontologies must not be ambiguous, and must make explicit the specification of a conceptualisation. This emphasis of ontologies as *'logical theories'* is also seen in [41], where this aspect of unambiguous interpretation is re-iterated. In contrast to this however, Guarino and Giaretta [50], weaken Gruber's definition, by stating an ontology is only a *partial* account of a conceptualisation.

'An ontology is a logical theory which gives an explicit, partial account of a conceptualization'

Here a more rigid definition of a conceptualisation is put forward, as an intensional semantic structure encoding implicit rules constraining the structure of a given reality. This definition, implies that an ontology cannot express a complete given conceptualisation but merely a part of it, allowing for different types of ontologies, one with a richer language, or one that relies on inference to express this conceptualisation. It is here that the element of logic is first stated, to determine that a conceptualisation is unambiguously defined, a concept that will become critical in ontology mapping and re-use.

Musen [94] refers to an ontology as a 'type of knowledge', introducing the importance of sharing and re-use of knowledge in the form of ontologies within the medical domain, and proposes the following ontology definition as:

"Formal descriptions of objects in the world, the properties of those objects, and the relationships among them".

here, Musen specifies the formal element of an ontology. Borst [21] directly furthers Gruber's definition to incorporate this commonality and the fact that an ontology needs to be shared:

'An ontology is a formal specification of a shared conceptualization'.

This definition maintains the formality of an ontology, as posed by Gruber, and by Borst, who reiterates the element of a shared conceptualisation, an idea routed in ontologies and their re-usability [94]. Borst states what makes an ontology useful is that it is formally defined, and thus reusable leading to the notion of *ontological commitment*. The concept of sharing and reuse have been key aspects of the definition of ontologies, thus encouraging the growth of the use of ontologies within the semantic web, where the focus is on creating a well defined and openly shared conceptualisation that is agreed upon and utilised by multiple systems.

Studer and collegues extend Borst's definition, adding the constraint of an unambiguous specification [120] stated by Guarino and Giaretta and Gruber, and reinforce the elements of formality, logic, and sharing:

'An ontology is a formal, explicit specification of a shared conceptualisation. Conceptualisation refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge that is, it is not private of some individual, but accepted by a group.

This definition highlights three important elements, which are seen across most of the definitions proposed, that an ontology should be *explicit*, *formal*, and *shared*.

Noy proposes a more user focused definition [99] of an ontology, that although centred in a designers perspective, still implies the formal, explicit and the use of a shared conceptualisation:

'An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.'

2.2 What is an Ontology?

It is clear from the variety of definitions presented previously in this chapter, that ontologies are not easily defined. Focusing on their use in AI and computer science the definitions are centred around a number of key elements. Therefore it is important to be clear on what are the main elements of an ontology, and their uses in relation to this work. Throughout the various definitions discussed above, there are a common set of elements that are explicitly stated:

i *Explicit*: Referring to the unambiguous specification of vocabulary;

ii *Formal*: Stating that the vocabulary is machine readable;

iii *Shared Conceptualisation*: A conceptualisation is shared and mutually agreed domain context.

From the above discussion of varying ontology definitions the following definition, this thesis modifies Studer's definition by emphasising interoperability.

Definition 1: *An ontology is a formal, explicit specification of a partially shared conceptualisation. In this context a* ***conceptualisation*** *refers to a domain that is defined by a vocabulary assigned.* ***Explicit*** *means that the concepts and restrictions detailing the vocabulary are formally defined, where* ***formal*** *is based in logics and machine readable.* ***Partially-shared*** *denotes that the ontology assigned to an agent, might not be shared in full, but only a fragment might be shared in a given interaction.*

This work maintains the importance of the defining features of an ontology including the explicit representation of terms, and the use of formality, however the notion of a *shared* conceptualisation is something that needs to be specifically noted with respects to the focus of this work.

As a result of the focus of this work which will be described in Part II, it is important to identify a more appropriate definition of a 'conceptualisation', as a conceptualisation which is partially shared. As the participating agents utilise partially shared ontologies, it is important to be aware that a conceptualisation of a world in an ontology by one agent, may never be seen in full by another agent.

In this thesis, the focus is less on the reuse of an ontology, rather addressing an ontology in its high-level meaning, as a set of well defined linked concepts referring to a domain. This work, does not seek to re-use ontologies, but rather, utilise their semantic structures in order to find consensus over concept meaning to generate an alignment through negotiation.

2.2.1 Conceptualisation and Ontological Commitment

Due to the subjective nature of ontological design, there are often multiple perspectives in modelling the knowledge of a domain. It is therefore unlikely that two agents would model the same domain using the same terms and level of detail. Given a particular domain, two system would have different conceptualisations of the same domain and represent the domain knowledge using different names and formalisms. Figure 2.1 visualises the relation of a given world or domain, the conceptualisation of this domain, and the ontological models representing the knowledge. It can be seen here that there is an intended model of the knowledge, to which is presented fully in ontology 2 and partially in ontology 1. This intended model could present a given task or a level of detail, in the knowledge represented by the ontology. As with their subjective nature, ontologies can represent varying levels of granularity when presenting a domain: this will be detailed in the next section referring to *ontology categorisation*. A conceptualisation, Figure 2.1, is a formal structure of a world or domain including the concepts and how those concepts representing the notions relevant for describing the domain relate to each other. Genesereth and Nilsson represent a conceptualisation as an extensional relational structure discussed in [60]. Here a conceptualisation is defined as a tuple *(D, R)* where D is the set of objects the knowledge represents, and R is a set of relations between the objects in D.

Also represented in this Figure 2.1 is the notion of *ontological commitment*. This is the agreement of the meaning of the vocabulary used in an ontology between two agents. In this work it is assumed that, each agent has a single assigned ontology, seen in Figure 2.1 as ontology 1 and ontology 2. This ontology will model a given conceptualisation of a given world and, may not be a full representation of an intended model, however it represents a single representation of a domain.

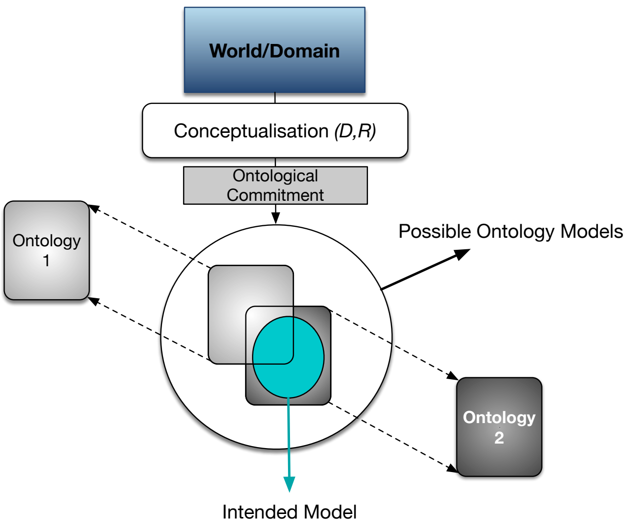


FIGURE 2.1: Ontology conceptualisation, based on represented model from *'What is an Ontology?'* [60]

2.2.2 Categorisation

Ontologies have long been used in agent systems to model a representation of knowledge, however the level of granularity of this knowledge is not uniform across ontologies. Ontologies can therefore be categorised depending on their level of specification, and how they model the knowledge they represent.

Guarino distinguishes ontologies by the concepts they model, as *Top-level*, *Domain or Task* ontologies and *Application* ontologies, a combination of domain *and* task ontologies [59]:

- **Top-level** ontologies describe very general concepts, independent of a particular problem or domain [59]. Top-level ontologies can be domain independent in their design such as models of time. These ontologies are domain independent and are not specific to a given task or problem, and describe concepts in a very broad and general way.

- **Task** ontologies are similar to domain ontologies, and describe concepts within a given domain in terms of a particular task, for example selling.

- **Application** ontologies are designed specifically both to model a domain *and* a given task. They often represents the concepts in terms of a particular task performed.

It is important to note that top-level ontologies are less specialised, and therefore are easier to share and reuse due to their more *'common sense'* content. In contrast to this, ontologies at the lower levels such as application ontologies, can only be shared if the other system accepts the models in the levels outlined above [97]. Ontologies can also be classified in relation to the granularity of the knowledge modelling. This categorisation includes *lightweight* and *heavyweight* ontologies. Lightweight ontologies are mainly represented with concept hierarchies including relations between them, whereas *heavyweight* ontologies add axioms and further constraints to the *lightweight* ontologies [54], Uschold [124] details four levels of ontology that depend on the level of formality in the modelling of their vocabulary. These levels are:

- **Highly-informal** ontologies represented in natural language, and thus are non machine-readable [120]. These ontologies, due to the nature of natural language, are unlikely to be explicit in the definition of their domain vocabulary.

- **Semi-informal** ontologies also represented in natural language, however the use of restrictions reduce the level of ambiguity.

- **Semi-formal** ontologies are machine readable, and expressed in an artificial formal language [124].

- **Rigorously-formal** are machine readable ontologies, formally and explicitly defining concepts and semantics including proofs of meaning in terms of soundness and completeness.

As outlined in this section, ontologies do not necessarily represent knowledge using a unique view. The domain they model can be at a very high and general level of detail, or be focused to a very specific task and domain.

2.3 Components of an Ontology

As previously discussed ontologies can fit into different categories regarding their level of detail in the content they model, and the level of formality of their represented language.

Regardless of these categories, ontologies are comprised of five main components. The following section details these main components that build up an ontology, and will be represented using an ontology modelled in the pizza domain, taken from [101].

As stated in the definitions of an ontology presented in Section 2.2 it is important that an ontology formally represents the vocabulary in a given domain by giving a specification of its meaning. This is done using five main components: concepts, individuals, relations, functions and axioms.

- **Concepts** *(C)* or also called classes, are central to most ontologies, and describe the objects in a given world, formally by using a term. Classes can be subdivided into subclasses which represent an entity that is more specific than that of its superclass. For example: in the food domain, classes include *Food*, *IceCream* and *Pizza* the latter two which are both subclasses of *Food*. This specification subsumption defines that all *Pizza* is a type of *Food*, and thus the child class, here *Pizza* would inherit constraints and properties from its parent class *Food*. This can be represented explicitly as:

  $$\texttt{Pizza} \sqsubseteq \texttt{Food}$$

  Other components of the food domain can also be modelled, such as:

  $$\texttt{IceCream} \sqsubseteq \texttt{Food}$$
  $$\texttt{Pizza} \sqsubseteq \neg\ \texttt{IceCream}$$
  $$\texttt{Pizza} \sqcap \texttt{IceCream} \sqsubseteq \bot$$

  Here the domain expresses that IceCream is a type of Food and Pizza is not IceCream and Pizza is disjoint to IceCream.

  Classes also formally detail the requirements and restrictions that must be adhered to for the membership to that class. A class restriction on the *Pizza* class, is that any concept in that class, must have some type of *PizzaBase* for it to be classed as a *Pizza*.

  $$\texttt{Pizza} \sqsubseteq \exists \texttt{hasBase.PizzaBase}$$

- **Relations** *(R)* or properties are the links between the concepts. Relations can have specific characteristics for example they can be symmetric such as:

  $$\texttt{Borough}\ \textit{hasNeighbour}\ \texttt{Borough}$$

  Relations can also be transitive such that:

  $$\texttt{Camden}\ \textit{isIn}\ \texttt{London}\ \text{and}$$
  $$\texttt{London}\ \textit{isIn}\ \texttt{England}\ \text{therefore}$$
  $$\texttt{Camden}\ \textit{isIn}\ \texttt{England}.$$

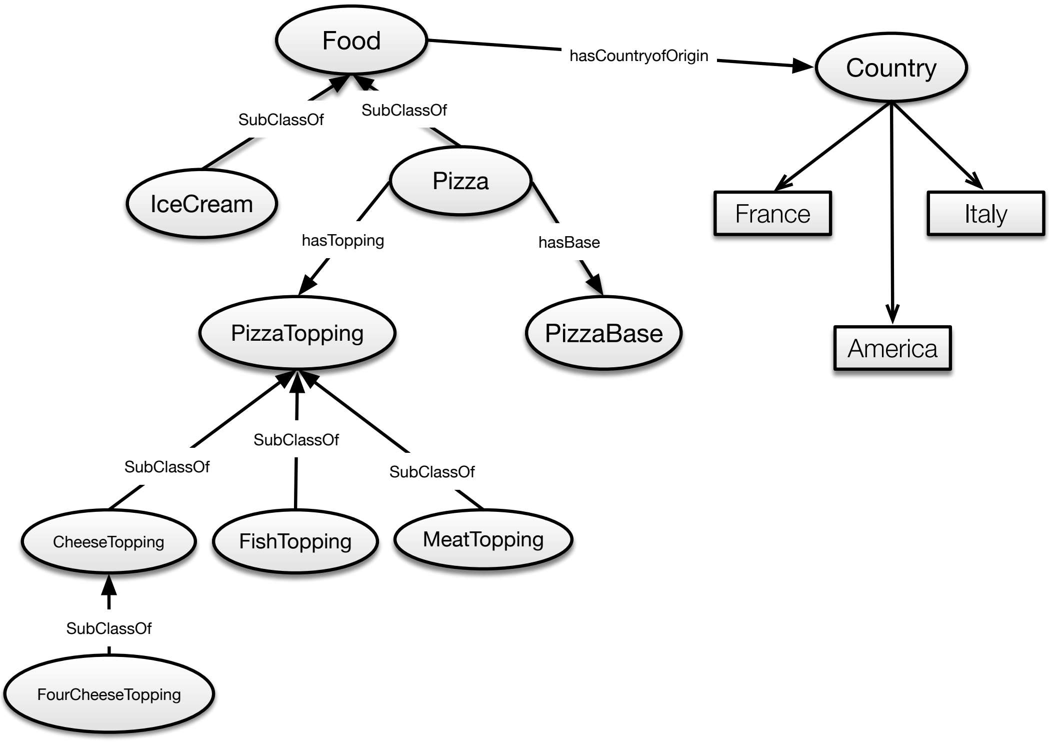


FIGURE 2.2: Ontology components from the pizza ontology, illustrating classes, relations and individuals. [101]

Relations all have a direction, the source of this relation is the domain, and the class it points to is its range. In the pizza domain, the relation *hasBase* is a relation between the class *Pizza* in the domain *(d)*, to *PizzaBase* in the range. This can be represented explicitly as:

$$\textit{(d)}\ \texttt{Pizza}\ \text{hasBase}\ \texttt{PizzaBase}$$
$$\textit{(r)}\ \texttt{PizzaBase}\ \text{isBaseOf}\ \texttt{Pizza},\ \text{isBaseOf} \equiv \text{inverseOf('hasBase')}$$
$$\text{Domain restriction: } \exists\text{hasBase.}\ \top \sqsubseteq \texttt{Pizza}$$
$$\text{Range restriction: } \top \sqsubseteq \forall\text{hasBase.PizzaBase}$$

- **Individuals** *(o)* are the instances or objects of a given world at the base level. Ontologies do not require individuals to be specified, however they allow a division between classes and instances of classes for a designed ontology. In the pizza domain *America*, *Italy*, *England*, *France* and *Germany* are examples of instances, specifically *America* is an instance of *Country.*

```
Country(America), Country(Italy), Country(England)
```

- **Functions** *(F)* are a particular case of relations which are defined on a set of concepts and are sub-relations from a given parent quality. These can relate an individual concept to a single value. Using the pizza domain, an example of a function is *Spiciness*, which is a functional relation of some *Topping* and a partitioned value from *hot*, *medium*, or *mild*.

$$\texttt{hasSpiciness} \in \texttt{Spiciness : (hot, medium, mild)}$$

- **Axioms** *(A)* are logical formula that can infer and define specific class restrictions not formally defined, that are always assumed to be true. They are commonly used to verify the correctness of the knowledge represented, through checking, and inference. An example of an axiom in the pizza domain, can be seen in *VegetarianPizza*, where it is specifically defined that no *FishToppings* or *MeatToppings* can be on a *VegetarianPizza*.

$$\texttt{VegetarianPizza} \sqsubseteq \neg\exists\texttt{FishTopping} \sqcup \neg\ \exists\texttt{MeatTopping}$$

2.4 Ontology Languages

Ontologies can be designed differently in respects to their knowledge content, however they must adhere to a specific language in which they are modelled in order to fulfil the *formality* element of the ontology definition. There are different languages used to formally specify and represent ontologies, which vary in expressivity. This section outlines the most common ontology representations, and summarise them in Table 2.1, from RDF/RDFS, and description logics, to the variants of OWL, a language designed specifically for the semantic web, in order to represent elaborate knowledge regarding 'things, groups of things, and relations between things' [82]. However in this work language assumptions are made, in which the agents use the same syntactic representation, therefore the type of language used is not important to this work, but it is assumed that the ontologies are represented in terms of concepts and relations.

- **RDF/RDFS**: *Resource Description Framework* designed by W3C for the semantic web, and is a standard modelling language for describing metadata. RDF models are based on triples, where an edge links the relationship between two given resources as ⟨s-p-o⟩ triple, where the *subject* and *object* are classes and the *property* is the relation linking them. These triples form a directed and labelled graph of the data represented. RDFS is the schema to RDF it allows entailment and inherits RDF syntax, and is fundamentally about describing classes of objects. For example using the pizza domain, the following RDF fragment defines the class *American*, and represents the relation that *American* is a subclass of *NamedPizza*, and is the domain of a relation *hasCountryOfOrigin*, where the range is *America*:

$$\langle \texttt{\#American}, \texttt{rdf:type}, \texttt{\#rdfs:Class} \rangle$$
$$\langle \texttt{\#American}, \texttt{rdfs:subClassOf}, \texttt{\#NamedPizza} \rangle$$
$$\langle \texttt{\#hasCountryOfOrigin}, \texttt{rdfs:domain}, \texttt{\#American} \rangle$$
$$\langle \texttt{\#hasCountryOfOrigin}, \texttt{rdfs:range}, \texttt{\#America} \rangle$$

- **DLs** *Description Logics*: are a formalism for representing knowledge that can also be used to model ontologies explicitly. Description logics can unambiguously model a network of atomic concepts, including classes and individuals *(classes)* and the atomic relationships between them *(roles)* [11] and then developed to model more complex relations and restrictions creating statements in a knowledge base.

  Within *DLs*, basic atomic concepts can be used to model more complex relations and restrictions, by using *concept* or *role* constructors and by adding restrictions on the latter. *DLs* syntax can model boolean constructors such as: conjunction, disjunction and negation.

  Using *DLs*, an ontology can be described by defining a *TBox* and an *ABox*, using terminological and assertional components. Firstly a *TBox*$(\mathcal{T})$ describe a knowledge base in terms of declaring its concepts and the properties between them. Terminological axioms can represent *equivalence* and *subsumption* relations.

  $$\texttt{CheesePizza} \sqsubseteq \texttt{Pizza}$$

  An *ABox*$(\mathcal{A})$ contains assertional knowledge about *TBox* concepts, and models specific properties of individuals, known as membership properties. Using the pizza domain for example, the following axiom represents the role assertions for a *Pizza*that is a *CheesyPizza* defined to be equivalent to a *Pizza* which must have a topping that is a *CheeseTopping*.

  $$\texttt{CheesyPizza} \equiv \texttt{Pizza} \sqcap \texttt{hasCheeseTopping.Pizza}$$

- **OWL**: RDF/RDFS is an efficient language for modelling in terms of triples, however does not have the ability to represent more advanced notions for example, a member of *Food* can not be both *Pizza* and *IceCream* at the same time. RDFS cannot represent these classes as *disjoint* from each other, meaning they do not share any of the same instances. The Web Ontology Language *OWL* is a modelling language for web ontologies [16] which uses URIs to denote an ontology and models it in terms of classes, individuals and the links between them, and can include given characteristics. The OWL ontology benefits from having DL as its main underpinning. OWL has three sublanguages: OWL Lite, OWL DL, OWL Full. OWL 2 extended OWL with a number of new features requested by users. The relation of these languages and sublanguages to each other, are represented in Figure 2.3.

  i **OWL Lite** is a sublanguage, and a more restrictive form of OWL DL, limiting the use of intersections and cardinalities and excludes the use of enumerated classes, unions and disjoint statements.

  ii **OWL DL** is centred in $\mathcal{SHOIN}$ an expressive form of description logics [11]. OWL DL requires classes, and individuals to be represented separately, but has good expressiveness in cardinality constraints and class descriptions.

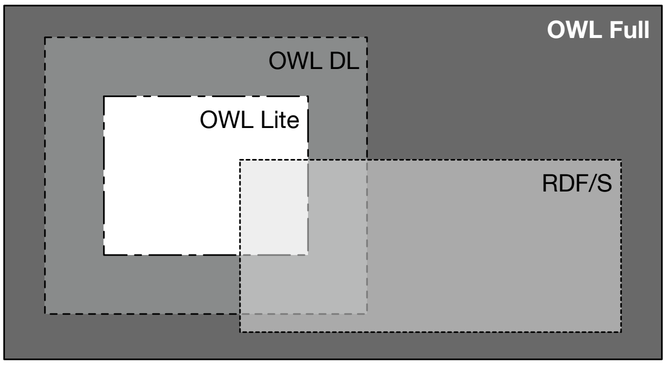


FIGURE 2.3: Language relations between RDF/S and OWL variants, based on W3C documentation. [82]

iii **OWL Full** is the most expressive form of OWL, and most closely linked with RDF/S by which OWL Full semantically extends the language. Although OWL Full is highly expressive, this characteristic means it is undecidable [12], thus there is no ability for complete or efficient reasoning. OWL Full due to its expressivity, distinctly has the ability to model a class as both a set of individuals or as an individual itself.

iv **OWL 2** [82] is an expressive extension to OWL, which adds increased expressivity. OWL 2 similar to OWL is centred on representing ontologies as RDF graphs. In comparison to OWL 1, OWL 2 adds expressive abilities on properties in terms of property characteristics and incompatibility, simple meta-modelling capabilities, extended annotation capabilities, and extends support for datatypes. OWL 2 also allows for negative property assertions, thus allowing for 'negative facts' asserting values that a given individual specifically does not have. OWL 2 also gives rise for self restriction allowing for the definition of an object class to relate to themselves though a property.

ObjectPropertyAssertion(a:likes a:Peter a:Peter)
ObjectHasSelf(a:likes) [82]

TABLE 2.1: Ontology Language Comparisons

Language:	RDF/S	DLs	OWL Lite	OWL DL	OWL Full	OWL 2
Universal Quantification	✓	✓	✓	✓	✓	✓
Cardinality Restrictions	✗	✓	`Binary`	✓	✓	✓
Self Assertion	✗	✗	✗	✗	✗	✓
Subsumption	✓	✓	✓	✓	✓	✓
Negation	✓	✓	✗	✗	✓	✓
Expressiveness	`Medium`	`Variable`	`Poor`	`Medium`	`High`	`High`
Compatibility with RDF	-	✓	✗	✗	✓	✓

2.5 Summary

An overview of the development of the area of ontologies has been presented beginning with their philosophical origins, to their use in artificial intelligence. Here the various definitions of an ontology in the artificial intelligence community have been examined in terms of bringing together the key elements within an ontology, which are highlighted in well used definitions. The components of an ontology are introduced in terms of the domain they represent. An overview into ontology languages, has been introduced in terms of how these languages differ from each other. This chapter, particularly focuses on the components and previous definitions of ontologies, which are key to the work presented, and has also introduced a definition of an ontology, which will be used throughout this thesis. Ontologies are commonly used within the agents community, and a well established area of research, is aligning ontologies, and the problems which arise due to the nature of heterogeneity in their design, will be addressed in Chapter 3.

In Summary of this Chapter:

- Defined and detailed the origin and development of the use of ontologies from their beginnings in philosophy to their use in computer science.
- Summarised the key elements taken from currently used definitions of an ontology.
- Detailed ontology commitment and conceptualisation, and the components making up an ontology in terms of entities and relations.
- Formal definition of an ontology as:
  A formal, explicit specification of a partially shared conceptualisation. In this context a *conceptualisation* refers to a given domain that is defined by a vocabulary assigned. *Explicit* means that the concepts and restrictions detailing the vocabulary are formally defined, where *formal* is based in logics and machine readable. *Partially-shared* denotes that the ontology assigned to an agent, might not be shared in full, as this work examines the concept of privacy in ontology sharing.
- Outlined and compared the varying ontology languages, used to model ontologies, and highlighted their differences.

Chapter 3

Literature Review: Ontology Alignment

Chapter Outline

'Do you wish me a good morning, or mean that it is a good morning whether I want it or not?' - J.R.R.Tolkien

Ontologies and the knowledge they represent of a world can differ in both language and structure due to both the nature of ontologies and the subjectivity of their design. The resulting heterogeneity means ontologies rarely match like for like across different autonomous systems, which poses a major challenge to ontology mapping. This challenge is known as 'the ontology alignment problem' and is centred around the heterogeneity of ontologies. Ontologies are heterogeneous if they contain different characteristics and as a result there exists many different forms of heterogeneity, each affecting ontology matching in different ways; these include semantic and syntactic heterogeneity. This chapter presents a review on the literature of ontology alignment: the ambiguity heterogeneity causes within the ontology alignment problem and summarises several current systems that tackle this subject using different approaches.

3.1 Ontology Alignments

An ontology is defined as a formal, explicit specification of a partially shared full conceptualisation (see Chapter 2). In this context a *conceptualisation* refers to a given domain that is defined by assigned vocabulary. *Explicit* means that the concepts and restrictions detailing the vocabulary are formally defined, where *formal* is based in logics and machine readable. *Partially-shared* denotes that the ontology assigned to an agent, might not be shared in full, as such this work examines the concept of privacy in ontology sharing.

A key element in this definition of an ontology is that of the conceptualisation, referring to the formal structure of a world or domain modelled by an ontology. This conceptualisation includes the concepts modelling a domain and how those concepts relate to each other, which is perceived by an agent though commitment to an ontology model.

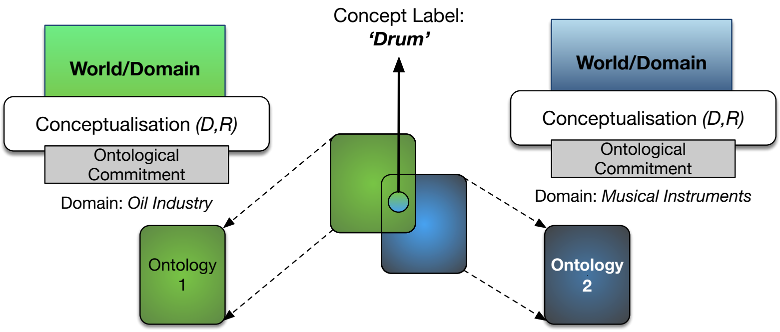


FIGURE 3.1: Introducing heterogeneity in interoperability.

This conceptualisation of a world is subjective in its nature and dependant on the requirements of the ontology, whether it is task or domain dependant, and therefore designed for a specific purpose in mind, or on the other hand independent of a particular task. This subjectivity of the vocabulary used in an ontology, therefore gives rise to the problem of interoperability between the agents over the meaning of their knowledge bases, due to *heterogeneity* seen in Figure 3.1. Here extending the model represented in Chapter 2 the varying conceptualisations adopted by the differing ontologies *ontology 1* (presents an oil industry domain) and *ontology 2* (presenting an instrument domain) introduce the notion of heterogeneity. Here it can be seen that both ontologies model the concept *'Drum'*, however the ontologies model different domain knowledge. Ontology 1 models an oil industry domain, and represents the concept *Drum* as an oil receptacle, and ontology 2, is a musical instrument domain that would model the concept *Drum* as a percussion instrument. If these two agents were to interoperate over their knowledge bases, they would be referring to two very different concepts.

For the two agents to accurately understand each other when interoperating over concepts, they must have common background knowledge to ensure that when they refer to a given concept, for example *Drum*, they are both describing a similar concept.

It is important that the common ground of interoperated meaning is highlighted by the agents and addressed, or they could produced an incorrect alignment across their ontologies defined as a mismatch (discussed in Section 3.2).

Ontology alignment is defined as the creation of a set of mappings between corresponding entities within a pair of ontologies [38]. An alignment, A, is a set of correspondences between ontologies and produces a set of correspondences in terms of a triple $\langle O_e,\, O'e',\, r\rangle$ and can be formally defined as:

Definition 2: *An Alignment A takes in two ontologies,* $O\{e\}$ *and* $O'\{e'\}$ *and produces a set of correspondence in terms of a triple* $\langle e,\, e',\, r,\, c\rangle$*, where* $e \in O$ *and* $e' \in O'$ *are the entities from the source ontology O and target ontologies* O' *respectively, and r is the relation between them, this can include equivalence, subsumption, disjointness etc,* (in this work, only equivalence relations are considered.) *Finally c is the confidence value attributed to the candidate mapping as a value in [0..1].*

This definition can be further specified in terms of the relations permissible between the ontology entities, restrictions on the relations themselves, and in terms of the level of completeness of an alignment, between a full or partial alignment over the ontologies included.

These alignments allow agents to identify mutual concepts between their ontologies and where relevant, enables the most efficient changes in terms of computational costs, re-structuring and re-designing their ontology in order for them to communicate effectively. Thus one approach of interoperation is to semantically map terms from one ontology to another, and generate an alignment across the knowledge bases. This alignment takes the two ontologies, seen in Figure 3.2 as O and O', and produces a semantic alignment A'. Alignments are a cost effective solution to interoperability as knowledge-based systems such as ontologies and services are expensive to build, test and maintain [58]. Heterogeneity is a critical problem found in mapping ontologies and in

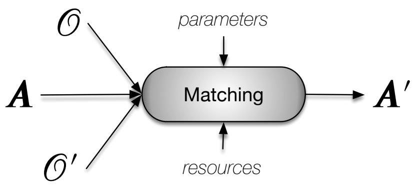


Figure 3.2: Matching process generating an alignment between two ontologies, taken from *Ontology Matching* [38]

the agreement of a meaning between the representation of different conceptualisations. Establishment of meaning is defined as *'semantic interoperability'* [80] and can affect how agents communicate accurately about similar concepts.

If different ontologies assigned to the agents are not accurately aligned, then one concept in an ontology my be matched to an unrelated concept in another. Previous research in ontological study has shown that alignments can vary in significance to a reference alignment. This mapping problem often arises when communication is based on different ontologies and is influenced by the heterogeneity between the given ontologies, both semantically and structurally.

In its highest level, an ontology alignment can be defined as *'any formal description of the (semantic) relationship between ontologies'* [138]. This alignment is an output of a given mapping process which takes in two or more ontologies, and is comprised of a set of correspondences, between an entity represented in one ontology and an entity represented in another. From this point onwards, this thesis refers to an alignment in its singular sense, i.e only including two inputted committed ontologies belonging to two independent agents, respectively, rather than addressing multi agent ontology mapping.

3.1.1 Types of relation

Within ontology alignments, the relations between the mappings from the source concept to the target concept can vary and commonly include equivalence ($\equiv$), subsumption ($\sqsubseteq$), and disjoint ($\perp$) relations. An equivalence ($\equiv$) mapping can be illustrated with the example: $\langle$*Song*, *song*, $\equiv\rangle$. This relation defines an equivalence mapping between the entity *Song* in a source ontology to the entity *song* in a target ontology. This relation forms one mapping in an alignment, and is usually supported by a confidence value represented as: $\langle$*Song*, *song*, $\equiv$, [0.8]$\rangle$. Subsumption relations $\sqsubseteq$, form mappings from one ontology to another, where a hyponym-hypernym relationship exists between the entity in the source ontology to the target ontology. This subsumption relation would be represented as: $\langle$*SportsCar*, *Car*, $\sqsubseteq$, [0.8]$\rangle$. A disjoint relation $\perp$, indicates the exclusion of one class from another, for example $\langle$*Song*, *Car*, $\perp$, [0.8]$\rangle$.

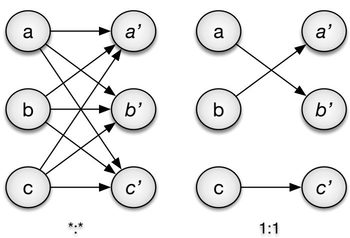


FIGURE 3.3: Multiplicity restrictions based on representation in [38]

Alongside the types of relation between entities from a source to a target ontology, there are important restrictions put on the number of allowed relations that are accepted into an alignment. This restriction, *multiplicity*, defines how many entities from a source ontology can correspond a single entity from the target ontology. These include, one to one mappings i.e. an injective mapping (1:1) and a many to many mapping (*:*) which are illustrated in Figure 3.3. Figure 3.3 illustrates the restriction between the number of relations between the concepts in one ontology $\langle$a, b, c$\rangle$ to the concepts in anther ontology $\langle a', b', c'\rangle$. Here on the left, it is permitted that multiple entities can correspond to a single entity within a mapping. In contrast to this, on the right a 1:1 restriction is illusrated, defining that only a single entity from one ontology can correspond to a single entity from another ontology.

Therefore these alignments can be either full alignments or partial alignments. Firstly a full alignment requires all the entities from one ontology to be mapped to all the entities of another. The aim of completeness would enforce a full alignment for all the entities in the ontologies, however this may compromise the accuracy of the mapping and result in mismatches which will be detailed in Section 3.2. In contrast to this correctness, only results in a partial alignment however the accuracy and confidence in the mappings can be much higher leading to fewer mismatches.

3.2 Example of Heterogeneity

Extending the introduction of heterogeneity seen in Figure 3.1, this section details examples of heterogeneity using two ontologies both representing music knowledge, but with differing domains and levels of granularity.

Figure 3.4 illustrates two ontology fragments represented and designed within a broad music domain. Ontology O is designed as a *music shop* focusing on retail including unit cost in the scope. Ontology O' models a *music encycopedia* including details of recording, writing, and the production elements of a musical release. A partial alignment is shown, made up of 5 represented mappings (m^n), from ontology O and O', where $m^1 = \langle$*Song*, *song*, $\equiv\rangle$, $m^2 = \langle$*Performer*, *Performer*, $\equiv\rangle$, $m^3 = \langle$*Band*, *band*, $\equiv\rangle$, $m^4 = \langle$*Artist*, *artist*, $\equiv\rangle$, $m^5 = \langle$*Record*, *record*, $\equiv\rangle$. This example shows a series of 1:1 mappings found in a *partial alignment* of the two ontologies inputted, as there remain entities that have not been mapped in this alignment such as *PriceofSong* from ontology O and *members* from ontology O'. If the process does not include mapping all the concepts in a given source ontology with any accepted relation from a target ontology then a partial alignment rather, than a full alignment is achieved.

Ontology 1 was modelled using the following assumptions: A *Song* has exactly one *SongPrice*, as with *Record*. A *Record* also has at least two *Songs* and a given *Song* can appear on one or more *Records*.

Ontology 2 was modelled using the following assumptions: the *members* can be in one or more *Bands*, and *Bands* must have two or more *members*. A *song_manifestation* has one or more *Composers* and a *Composer* can compose one or more *song_manifestation*. A *song_manifestation* also has one or more *performers* and a *performer* can perform on one or more *song_manifestation*. *Song* has one or more *artist* (allowing for collaborations), and an *artist* can sing one or more songs. The designing of a separate *song_manifestation* to the concept *song*, allows for the modelling of both original songs and covers. In this case the *song* would be the version performed by either an original *artist* or a covering *artist*, and the *song_manifestation* is the original song with the composer as the musician writing the song. The *artist* definition is a single individual performing on a song e.g. Adele. A *song* is on one or more records and a *record* has one or more *songs*.

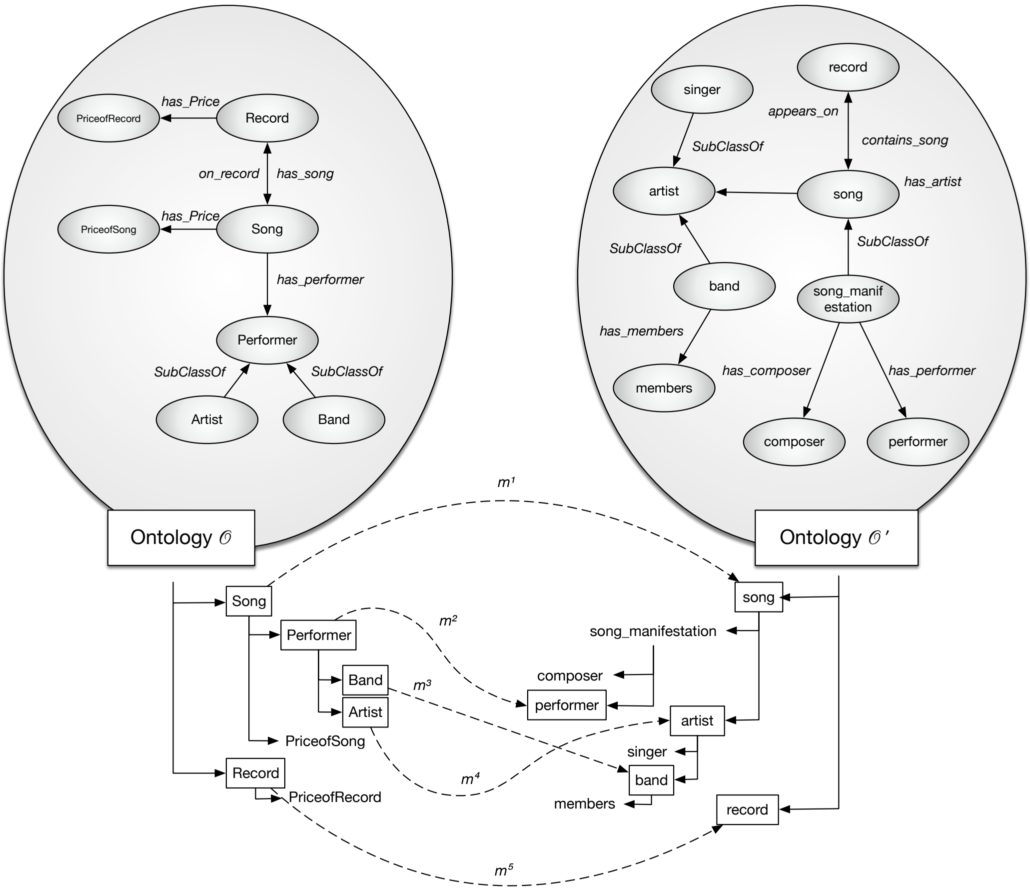


FIGURE 3.4: Alignment examples

With these defined assumptions, heterogeneity within the mappings can be seen. Using the concept *song* from ontology O' potential heterogeneity between the ontologies can be examined, when mapping it to the similar concept in ontology O.

Take the modelling in ontology O, where a '*Song* has a *Performer*' i.e. the group or individual performing a song, for the record to be purchased. In ontology O' the relation '*song* has an *artist*' models the same relation '*Song* has a *Performer*' in ontology O. The concept *performer* in O' is modelled concerning a *song_manifestation*.

The heterogeneity between the two ontologies is a result of the differing tasks they have been designed for, and are at a syntactic and semantic level. Ontology O as a shop models cost and price, and ontology O' does not include these, and instead represents details of a release. As a result of this heterogeneity a better mapping that takes into account this syntactic heterogeneity in the ontologies, can be represented in the mapping m^2= ⟨*Performer*, *artist*, ≡⟩ or m^1= ⟨*Song*, *song_manifestation*, ≡⟩.

When generating a full or partial alignment with this example given a restriction of 1:1 mapping enforced, the mapping accuracy would decrease. This could force mismatches *in leu* of completeness in order to generate a full alignment. Often in this case, the supporting confidence values for mappings in the alignment are lower.

3.3 The Alignment Problem

Ontologies are designed for specific tasks or for a particular domain and therefore can vary greatly in terms of structure, granularity and nature of design. This subjectivity causes problems regarding the meaning of terms when the two agents try to communicate. Take for example two individuals, each with a different mother tongue. They are gathered together to discuss a concept label they both identify as ‘sport’, however they both have very different languages describing the given domain, and the structure in which they categorise the subject. This is an example of heterogeneity, and can give rise to mismatched concepts if not properly resolved.

In order for agents to successfully communicate over their committed ontologies, the knowledge bases are input into a mapping process. This mapping process allows the concepts in the ontologies to be mapped to a seemingly corresponding concept in the other ontology in order to generate an alignment over the two.

Mapping ontologies is not always an elementary process and is often complicated by heterogeneity giving rise to *the ontology matching problem*, where agents need to communicate with each other, yet commit to different ontologies of varying definitions. This task is exasperated by semantic or syntactic heterogeneity in the design of ontologies, which makes the mapping process difficult, thus this complication needs to be taken into account and resolved. The categorisation of these mismatches is not a black and white process, there can be a lot of crossover between the resulting mismatches as a result of varying types of heterogeneity. A categorisation of these mismatches is presented taken from Visser [130] and further developed by Klein [76], where a similarity of structural or *syntactic* and *semantic* divide can be seen across the two.

3.3.1 Syntactic Mismatches

Visser [130] outlines two types of syntactic as *paradigm* and *language* heterogeneity:

- **Paradigm** heterogeneity is categorised by Visser as syntactic and occurs when different paradigms are used to model the knowledge;
- **Language** heterogeneity is syntactic and occurs when the two ontologies are represented in different ontological languages.

Klein develops Visser categories specifically in terms of language level and ontology level mismatches, aligning to syntactic and semantic heterogeneity respectively. Firstly language level mismatches occur when two participating ontologies are represented differently in terms of design language and logical restrictions, as described in Chapter 2.

This work is focussing on matching ontological concepts and resolving semantics, rather than ontology languages, therefore semantic mismatches are detailed below.

3.3.2 Semantic Mismatches

Visser [130] outlines two types of semantic heterogeneity firstly as *ontology*, and *content* heterogeneity and details ontology heterogeneity in terms of *conceptualisation* and *explication* mismatches:

- **Ontology** heterogeneity occurs when the two participating ontologies make different assumptions about their given domains; *Conceptualisation* mismatches occur when different conceptualisations (concerning the definitions of a conceptualisation detailed in Chapter 2) are used to represent a given domain, in terms of the concepts used, and how they are related within the knowledge base.
  Explication mismatches occur when there are differences in how the two conceptualisations are specified in terms of the concepts. This is detailed by Visser in terms of a combination of mismatches over the term T being defined, the factors which define it D and a concept description of C. Using the example detailed on the music encyclopedia ontology O', a concept description would be 'a singer is an artist who performs solo' where *singer* is the T being defined, and are explicated as $singer(X) \leftarrow$ artist(X) $\wedge$ solo, *artist* and *solo* are the D definiens used to describe the term T.

- **Content** heterogeneity is a semantic heterogeneity where the ontologies represent different knowledge.

Klein further develops Visser's categorisation of semantic heterogeneity, which are subdivided by Klein into the following types of mismatches:

- **Conceptualisation** mismatches are semantic mismatches that can be found, as a result of conceptualisation differences between the ontologies; *Scope* mismatches are seen when instances belonging to classes representing the same concept are different. In the given example here, solo artists in O are named in the *Artist* class, however in O' solo artists are named in the *singer* class. *Model coverage and granularity* mismatches are seen over the differences in the range of the domain that is covered by the ontology. The model coverage and granularity mismatch can be seen when one given ontology has a differing level of detail of a given domain than that of a second ontology. This can be seen in the given example in Figure 3.4 with the concepts $SongPrice$ and $RecordPrice$ both of which are represented in the given ontology O but not in O' due to the level of detail on the specific areas of the music domain they model.

- **Terminological** mismatches relate to the specific term used to label a given class within an domain. Firstly *synonym* mismatches are when the two given ontologies use different terms in order to define the same concept, for example in the terms *football* and *soccer*. In this case they are both referring to the same idea but with different terminology. Within this given example it can be inferred that from the modelling $Performer \equiv artist$, thus illistrating a synonym mismatch at m^4 between the terms *Artist* and *artist*. Secondly *homonym* mismatches are when the same label is used to represent different concepts from different domains, for example within in a music domain the term *Drum* referring to a musical instrument would have a different meaning than the use of *Drum* in an oil company domain referring to a barrel.

- **Explication** mismatches are taken from Visser categorisation of ontology level mismatches where the style of modelling differs between the two ontologies. These mismatches occur when firstly, different *paradigms* may be used within the two ontologies. Secondly *concept description* mismatches are where the same concept can be modelled in different ways such as using a qualifying attribute or by generating a separate class.

3.3.3 Current Alignment Systems

Many of the current approaches to ontology matching take in two (or more) ontologies, share these ontologies in full, and proceed to generate an alignment semantically, syntactically, or in some cases both. Most systems conduct a level of preprocessing on the input ontologies. These systems for matching are primarily automatic however there are some semi automatic systems which require user input.

In addition to semantic or syntactic matching methods, current state of the art for addressing heterogeneity involves a number of different approaches such as using background knowledge of a given concept in an ontology such as using external knowledge like WordNet [47] to provide semantic meaning [32, 45, 68] or using the structural locality of a class to provide a meaning through context [33].

A second addition to ontology matching is a focused task based approach, this approach introduces the notion of having an understanding of the task for which an alignment over two ontologies is required [79, 129].

Knowledge-based systems such as ontologies and services are expensive to build, test, and maintain [58], therefore it is important that if ontologies are to be mapped, the correct information is required to generate a meaningful alignment. The vast majority of current ontology alignment systems address the problem from a different focus. The focus of the majority of systems is where the ontology is shared in full prior to an alignment being generated. A successful alignment system is key in generating this meaningful alignment, and it is important that the alignment system is focused on the appropriate specification of the alignment task at hand. For example the size of ontology, computational costs, run time, or particular focus in generating a mapping.

This section outlines seven current commonly used ontology matching systems, which have been taken from [116] representing systems with different matching processes, including *FalconAO* [63], *Quick Ontology Matching* (QOM) [36], *S-Match* [52]. Also detailed are three systems, which have been evaluated on the Ontology Alignment Evaluation Initiative (OAEI) [67] and consistently generated good ontology mapping including: *XMAP* [32], *LogMap* [68], *Agreement Maker Lite* (AML) [45], *CROSI* [72].

AML: *AgreementMakerLite* [45] illustrated in Figure 3.5 is an ontology matching mechanism that is based on the AgreementMaker System, and focuses on dealing with large ontologies whilst maintaining computational efficiency, and continuously performs well within OAEI evaluations. AML is evaluated using large biomedical ontologies on the OAEI, and is comprised of two main modules: an ontology loader for inputting ontologies including those used for background knowledge and generating ontology objects and an ontology matcher which uses matching algorithms to align the ontologies, which can be illustrated in Figure 3.5 outlining the basic workflow for the system.

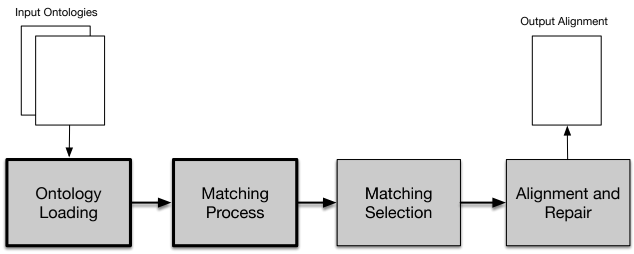


FIGURE 3.5: Generalised AgreementMakerLite Architecture

- *The AML mapping process:* AML is designed primarily with the idea of flexibility and extendability and can be used with most matching algorithms, and generates mappings on a multi step process, divided into *primary matchers* using 'HashMaps' and *secondary matchers* using term by term comparisons between concepts in the two inputed ontologies. Key elements of the mapping process within AML introduces a novel background knowledge mapping and if this is accepted above a given threshold this mapping can extend the *lexicon.* Due to AML primarily being used to map biomedical ontologies these background sources are mostly specifically designed to address specific medical terms, however WordNet is used for more basic English language vocabularies. Within the OAEI 2015 submission [43] the system used a number of steps: baseline matching, background matching, string and word matching, structural matching and finally property matching. These matchers then generate a set of mappings which are then stored in the framework's *alignment store.*

- *The AML selection process:* These mappings are selected from the multi step mapping process based on the *lexicon* data structure of the system. This lexicon holds the labels and synonyms of ontological terms where the matches are scored by a ranking system which assigned weights to a table of class names and synonyms taken from the ontology [44]. AML uses a 'greedy' mapping selection, divided into three selection criteria [43] depending on the size of the ontologies that are being used. Firstly *strict* selection forces a 1:1 alignment, so no mappings can include the same classes or properties, both *permissive* and *hybrid* selection allows for the same classes or properties to be mapped again, depending on the similarity score.

- *The AML repair process:* AML uses the alignment repair method generated in [111], which aims to repair mappings generated by an alignment which include classes seen as unsatisfiable, that are incoherent with the restrictions on semantics of the imput ontologies.

LogMap: illustrated in Figure 3.6 is an ontology matching system that has been designed to work with semantically rich ontologies [68], consisting of a large number of classes in the input ontology hierarchy. LogMap generates a set of initial mappings, and focuses on an iterative process of repair. Similar to AML, LogMap is evaluated using medical ontologies, due to their extensive size, and has performed well in OAEI evaluations.

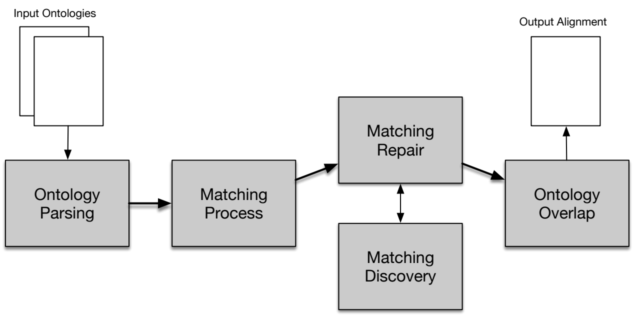


FIGURE 3.6: LogMap Architecture

- *The LogMap ontology input:* LogMap has a two step parsing process, on the lexical and structural elements of the input ontologies. Firstly it generates a lexical index of the class labels and their variations using external systems such as WordNet a step similar to AML and its use of background sources. Secondly the structural indexing, extends the ontologies hierarchy to add further information particularly regarding the disjoint classes of an ontology.

- *The LogMap matching process:* LogMap generates an initial set of exact string mappings from the input ontologies, which are then used as an 'anchor point' for the repair and discovery.

- *The LogMap matching repair and discovery:* LogMap utilises an iterative process, which takes in initial mappings that have been found and uses a combination of matching repair and matching discovery, in order to generate a more accurate candidate alignment. This processes and corrects those mappings which may be unsatisfactory in the semantic hierarchy, and the class restrictions of the input ontologies. Firstly in the matching repair, LogMap identifies the logically unsatisfactory mappings, using the Dowling-Gallier algorithm [34]. If a mapping is identified as unsatisfactory it is repaired using a 'greedy' process, by using the propositional representations of the individual ontology's classes in the unsatisfactory mappings, and repairs using a top down strategy in the hierarchy. Secondly the mapping discovery is done on the basis of neighbourhood similarity, where the semantically related class labels in one ontology are matched to those in the other, from a correct matching. LogMap iterates through the repair and discovery process until no further expansions of the context can be made in the matching discovery step.

- *LogMap overlap and output:* The last step of the LogMap matching process assesses fragments of the ontologies, that may contain additional mappings the system may have overlooked, in order to finalise an output alignment between the two ontologies.

Xmap: illustrated in Figure 3.7 is an ontology matching system that focuses on an aggregate alignment from lexical and structural class matchings [32], and are also evaluated using large biomedical ontologies, similar to that of the other systems.

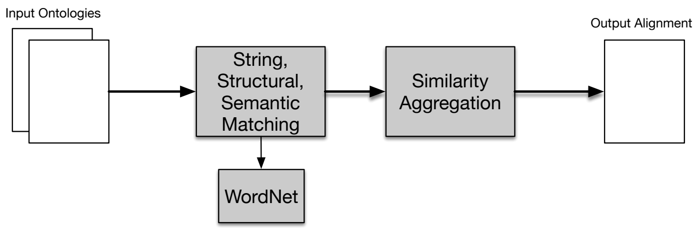


Figure 3.7: Xmap Architecture

- *The XMap matching process:* Xmap matches ontology concepts on two levels lexical and structural, and is furthered in Xmap++ to use external knowledge to include semantic matchings using external knowledge such as WordNet, specifically their *sysnets*, to provide 'local context' [33] giving meaning to the concept to map. The lexical matching is conducted by string matching algorithms on the class labels of the input ontologies. The structural similarity is done by comparing the cardinality restrictions, and hierarchy given to the classes. These two matches are then aggregated and if considered similar over a predefined threshold thy are are accepted to an alignment.

CROSI: is a semantic based mapping system, which uses name matches and semantic matchers to align two ontologies. CROSI illustrated in Figure 3.8 consists of four main modules including: feature generation, feature selection and processing, aggregation and evaluation [72] and uses aggregated similarities to iterate through the different stages, using multiple matchers independently. CROSI takes in OWL ontologies and feeds them into a feature selection module where the matching process begins.

- *The CROSI matching process:* CROSI uses a variety of different independent mapping techniques, including name matchers and semantic matchers, and takes into consideration structural and intension awareness. Within the various name matchers used by the system it incorporates the use of string distance as well as edit distance; both of which are seen in many of the other systems using lexical similarity. Within CROSI the structural awareness of the semantic matchers allows the loop to traverse class hierarchies and generate similarities throughout the sub-class relationships. Alongside this, intension awareness takes into account the different definitions of the classes matching property names, domains and ranges, thus giving a perfect match. CROSI also takes into consideration external matches in the mapping process, all of which provide further mapping options within the system. These given mappings are then passed into a similarity aggregator as seen in Figure 3.8, taken from Kalfoglou et al [72].

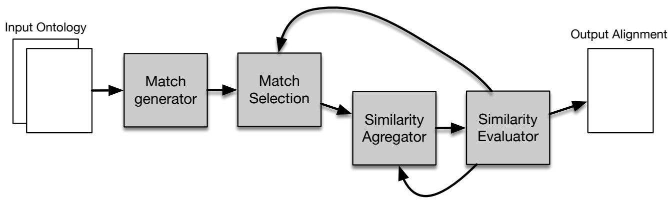


Figure 3.8: CROSI Architecture

- *The CROSI selection process:* CROSI selects the proposed mappings using multiple loops, in which the similarity is evaluated by users or supervised learners to begin intreating though the various stages of the mapping method process in the feature selection.The CROSI system is highly adaptable with its 'plug and play' design of using external mapping process, thus performs significantly differently using the various different combinations of these external modules.

  Falcon-AO: is an automatic system for mapping ontologies designed to use linguistic (LMO) and graph based (GMO) matching in an integrated process in order to produce an alignment of mappings between the ontologies. Falcon-AO in Figure 3.9 takes in two given ontologies, and through a parser based on Jena, removes redundant and inferred axioms as a result of heterogeneity.

- *The Falcon-AO matching process:* The linguistic matching (LMO) is conducted through lexical comparison which calculates the edit distance between two entities and generates a string similarity. Alongside the LMO, the system uses graph

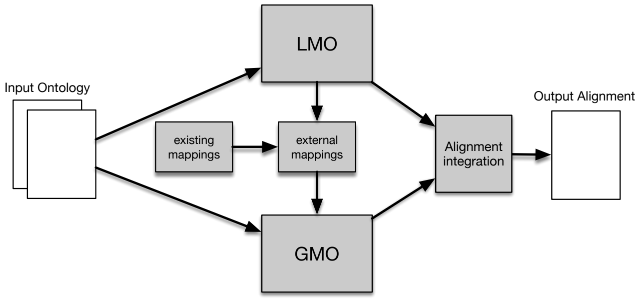


Figure 3.9: Falcon-AO Architecture

matching done through developing bipartite graphs. In the GMO [62] Falcon-AO uses the bipartite graphs to represent the ontologies, which are then co-ordinated using given rules in order to eliminate heterogeneity within the given ontologies. These rules are discarding, merging, inferring and list. GMO then determines external entities for the given ontologies and develops an external similarity matrix using WordNet. When the matrix is developed the system is run iteratively updating the equations until a pre-defined precision is reached giving 1:1 mappings. Finally the system outputs additional matched pairs between the ontologies based on structural similarity. GMO always attempts to find all the possible matched entity pairs

- *The Falcon-AO selection process* Falcon-AO uses both the GMO and LMO alignments and integrates them in order to produce final mappings between the ontologies, which are selected using the following rules: All alignments generated by LMO are accepted as accurate in terms of linguistic similarity. Within GMO, only the alignments with the high similarity are accepted into the alignment integration. When linguistic comparability is high and the structural comparability is low, only GMO alignments with high similarity are accepted by the system. And finally, if the linguistic comparability is low, all the GMO alignments are accepted as it is determined that there is insufficient information to measure the alignments. As discussed by Hu et al [63] overall Falcon-AO has been found to perform best when the ontologies have little lexical similarity, but high structural comparability.

S-match: primarily a semantic matching system platform [52], that takes tree like structures and transforms them into lightweight ontologies in order to establish semantic correspondences using semantic matching, minimal semantic matching and structural preserving semantic matching (SPSM). In comparison to Falcon-AO and OMEN, S-match does not perform graph based matching, but instead uses graph structures in the form of trees as an input format in order to process the given ontologies.

Illustrated in Figure 3.10 taken from Giunchiglia et al [51] S-match works over four main steps. Initially S-match loads in given ontologies and feeds them through a prepossessing, which contains the translation of natural language metadata in order to extract meaning and secondly a classifier step which constructs concepts and outputs these as description logic formula, in order to reduce heterogeneity.

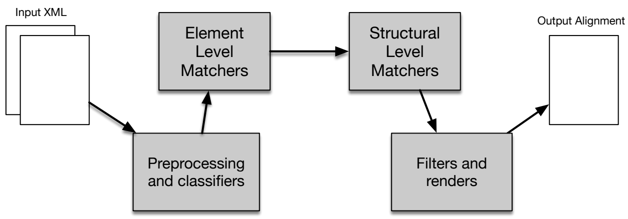


FIGURE 3.10: S-Match Architecture

- *The S-match matching method:* Step three of the system is the initial step of the matching process, which focuses on computing relations between the labels of nodes, in comparison to step four which computes the semantic relations between the concept of nodes. In step three semantic and syntactic matchers are used to compute the relations between the labels of the two given input trees. This is done semantically through linguistic matching and if there are no given semantics, then syntactically using N-gram and edit distance: a step which can be closely compared to the LMO process in Falcon-AO. This third step outputs a matrix of the relations of all the concepts in the nodes of the given ontologies. In step four, the process computes semantic relations between the concepts at the nodes of the given trees. Using the input from the previous step, the problem focuses on a propositional satisfiability between the pairs of nodes from the initial input ontologies.

- *The S-match mapping selection and output:* S-match uses filters and renderers in order to select the mappings for output. This is done using Semantic mapping techniques; minimal semantic mapping and SPSM. Minimal semantic matching provides the minimal set of unique matches between two ontologies, and SPSM computes the set of mappings whilst preserving the structural properties. S-match then outputs a set of mapping elements between the nodes from the two inputs. S-match is seen as a good matching system as it is customisable and can be used for many different applications.

QOM: a semi automatic mapping system, which operates using a trade off between effectiveness and efficiency. QOM is not a graph matching system, and uses no graph matching in the process of developing alignments between the ontologies. QOM uses string based similarity measures, comparable to those seen in CROSI and LMO of Falcon-AO, and also in the selection step of S-match [36]. QOM is comprised of six key stages seen in Figure 3.11, as part of the input and preprocessing, QOM initially inputs the ontologies and processes them in RDFS format, which allows them to be used in the four similarity calculations.

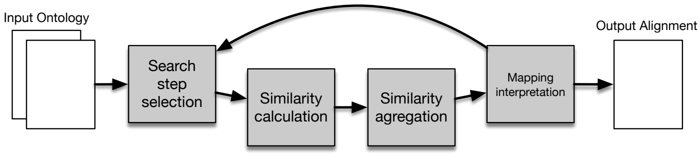


FIGURE 3.11: QOM Architecture

- *The QOM matching process:* In a search step selection, the QOM compares both the ontologies in full, finding any pairings and treating them as candidate mappings. This is followed by similarity computation, determining the similarity values of the candidate mappings found previously, which is done using four similarity measures: *object equality, explicit equality, string similarity and SimSet.* These similarities are then outputted and aggregated into a single similarity value.

- *The QOM selection process:* QOM uses interpretation, which takes the aggregated values to derive mappings. It does this by discarding mappings that don't meet the similarity threshold definition. QOM iterates this process until there are no new 1:1 mappings proposed.

Overall these seven ontology matching systems all work well with large ontologies, and have consistently performed well under the OAEI. They highlight the importance of scaleability for matching systems and focus on addressing heterogeneity within ontology design in order to generate meaningful alignments. These systems all differ from each other while covering similar factors in the process, illustrating that successful alignments can be generated in a variety of ways (e.g. string matching, and graph matching) and still produce a meaningful alignment across two ontologies. An overview of these systems can be seen in Table 3.1, which shows the comparisons of the systems in terms of their input and output, and mapping processes in terms of structural and string based methods.

System:	XMAP	LogMap	AML	CROSI	FalconAO	QOM	S-Match
Input size	Large	Large	Large	Large	Large	Lightweight	Lightweight
External inputs	✓	✓	✓	✓	✓	✓	✓
String matching	✓	✓	✓	✗	✓	✓	✓
Structural matching	✓	✗	✓	✓	✓	✗	✓

TABLE 3.1: Mapping Systems process overview

3.4 Summary

This chapter presented a detailed overview of ontology alignments including a formal definition, and emphasises the importance of agent interoperability, primarily in that successful communication relies on the agents having a understanding of the conceptualisation of a domain they are both referring to, within their respective ontologies. The focus of addressing this heterogeneity is detailed within ontology alignments, which is not always a straight forward process. The different forms of semantic and syntactic heterogeneity that occurs in mismatched mappings have been described. An example of this heterogeneity has been presented in a music domain. Finally current alignment systems have been compared, which have all produced successful alignments evaluated by the OAEI.

In Summary of this Chapter:

- Defined and detailed ontology alignments and the need for agents to establish meaning of their opposing agent's knowledge including the following formal definition of an alignment A as taking in two ontologies, O and O' and produces a set of correspondence in terms of a triple $\langle e_o,\ e'_o,\ r,\ c\rangle$, where e_o and e'_o are the entities from the source and target ontologies respectively, and r is the relation between them, this can include equivalence, subsumption, disjointness etc, *(In this work, only equivalence relations are used.)* Finally c is the confidence value attributed to the candidate mapping.

- Detailed the issues which arise in the heterogeneity of ontologies when generating ontology alignments.

- Outlined and compared current ontology matching systems used in current state of the art.

Chapter 4

Literature Review: Dialogue

Chapter Outline

'Talk is free but the wise man chooses when to spend his words.' - Neil Gaiman

Communication is an integral part of human cognitive interaction. In order for it to be successful, participants must co-ordinate over a joint task through which communication achieves a solution. This chapter presents communication from a cognitive perspective between humans interacting in order to complete a particular task. The main principles identified as integral in human communication have been addressed within communicative interaction between multi-agent systems thus introducing the notion of a dialogue. A dialogue is defined and detailed in terms of the components and rules, including Grice's co-operative maxims and Pask's rules of a strict conversation. It is through the use of a dialogue where the agents have a choice of defined moves built on the notion of ***Speech Acts*** *where the agents co-operate to complete a given task. Lastly, this chapter introduces the notion of ontology alignment generation addressed as a form of dialogical interaction and details general dialogues and the properties that are required for them to be developed successfully.*

4.1 Communication and Interaction

'Almost all animate organisms communicate' [19] however, the form and nature of this communication can vary in complexity. These communications can be through chemical releases seen in ants, sonar signals omitted by dolphins or indeed the use of spoken language, one of the varying elements of communication methods used by humans. These varying methods of human communication include visual stimuli in terms of physical messages such as sign language, body language or auditory stimuli including spoken language, all of which are designed to pass on messages from one person to another. These messages are usually over a given task, for example Figure 4.1 illustrates a child learning a new word from a parent. A parent in this case may use a physical stimuli in pointing to a 'thing' in a world and identify it with a unique name using spoken language, e.g. pointing to a ball, and saying to the child: 'ball'. The child can then learn that the identified object is called a 'ball'. This example, although using multiple sensory inputs, is related to word matching and can be seen as language acquisition and second language learning.

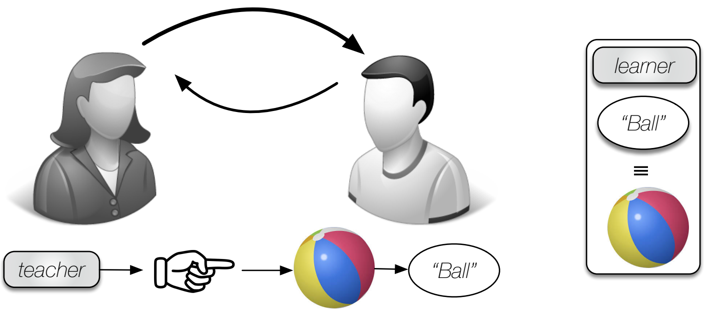


FIGURE 4.1: Human communication example showing object identification.

Human communication has been studied across numerous fields including philosophy and psychology, which has developed and inspired areas of research within artificial intelligence and multi-agent systems. Due to its multi disciplinary nature, human communication covers a large subject area too broad to detail for the purposes of this thesis. Although extensive, some of the principles of communication can be utilised in terms of how participants interact within a conversation in order to pass appropriate messages successfully, thus achieving a goal. The study of human conversation has philosophical origins beginning with Plato and the Socratic dialogues [71]. These dialogues first defined what constituted a conversation concerning the participants and their assigned roles introducing the need for co-operation, a key principle representing the successes of human conversation.

Language in the psychological field covers the varying ways humans communicate and can be studied as language acquisition in the formulating and understanding of spoken words. Varying approaches to researching human communication in psychology highlight the division between understanding an input and matching that to a given object or word. Jackendoff's language system [133] presents language learning as a *combinatorial cognitive process* highlighting the difference between sound and meaning. The importance of the distinction between understanding the input and understanding the meaning of the word is clear and mirrored in language acquisition in children. Noted by Clark, when children are learning a first language they can understand more than they can produce [27]. In first language acquisition, the importance of communication based learning of the language provides the learner with the understanding of specific items and the concept of turn taking is learned, which aids interaction [132].

Similarly with second language learning, humans have to match the word spoken using all inputs available in order to pair the word with something semantically similar to what they already know. A knowledge base or (first language) is known to a 'listener' and a different knowledge base (the new language) is known to an 'expert'. When the expert says a word the listener needs to find a similar word in their knowledge base. This process is similar to that of ontology matching and finding a meaningful semantic match between two terms as detailed in Chapter 3.

Language typically evolves in terms of what is *needed* and also from the point of view of the participant's *abilities* for example, in order to put out a fire you firstly need a fire. This two part nature of language evolution has provided humans with an extensively expressive communication system [26] in which they are able to assign meaning to complex stories such as the plot in a vast soap opera to the meaning of basic language through conversation. The development of a conversation using spoken communication is not uniquely human but with vast human communication capabilities, humans have developed conversations in which step by step utterances are made in order to carry out certain *joint actions*. This joint action is *'an action in which several agents co-operate to achieve a common goal'* [28]. For example a joint action between two participants could be performing a musical duet or indeed communicating a process for a team to pass a ball between them, and move forward on a pitch in order to complete a given goal of scoring. The success of a joint action vitally requires the co-operation of the participating agents in order for a goal to be achieved. Human conversations can be composed of verbal and non-verbal communications between participants, however the notion of context, meaning and co-operation remain clear factors to be understood. This allows a more formal view of a conversation where the *roles* of the participants involved are defined, along with the *language* used, the *goal* which is under 'conversation' and finally the *rules* to which the participants adhere to in order for these factors to be upheld.

As with humans in natural conversation, multi-agent systems still have to overcome one of the fundamental problems of language, that of understanding and meaning. The recent rise of cognitive computing reiterates the notion posed by Licklider in the rise of *human-computer symbiosis* [74] where computers and man are coupled tightly together, has given the opportunity to develop computer systems centred in human cognitive processes. Human communication made translating the capabilities gained from negotiation and co-operation, to the use in multi-agent systems greatly improve the flexibility and capabilities of these systems, for example in database integration, co-ordination of robots and dialogue systems [109].

These principles of human conversation have inspired research within artificial intelligence *(AI)* and multi-agent systems in the study of communication and interaction between agents [40, 118]. A key difference between human and agent communication is the restrictive nature of finite agent languages.

In multi-agent systems conversations can be defined in terms of a dialogue. Given multiple agents a dialogue is a series of arguments or statements made by the participating players or agents taking turns, where the first argument would be posed by $agent_1$ then a second by $agent_2$, until the goal of the game is complete. In the case of multi move dialogues, an agent would make all the moves they require, then followed by a second agent with all their moves until the game is complete. Wooldridge defines a dialogue formally in terms of moves and players [135], representing the two as a pair including the agent posing an argument and the argument itself.

Similar to language learning in humans, autonomous multi-agent systems can learn new terms through interaction with a 'teacher' agent. This can be seen in the *Naming Game* [118] where robots learn to identify objects with a given term from an independent 'teacher' robot, without prior knowledge regarding the objects or terms. Luc Steels contextualises the problem of ambiguity in language learning in the *grounded naming game*.

The *Naming Game*, represented in Figure 4.2 takes place between a set of agents, called the population (where population is denoted as P such that $P= \{ a_1..a_n\}$) where the roles of speaker and learner are randomly assigned. The agents have the same context of objects $o_i \in O$ (a set of objects), out of which the speaker internally selects one object referred to as the *intended topic*, and the game begins. In the first move of the game the speaker utters the name for the intended object. The learner then interprets the utterance and points to the object they perceive as the one in the utterance, this is the *perceived topic*. The speaker then agrees or disagrees with the learner if they have selected the correct perceived object. If incorrect, the speaker points again to the intended object and finally the speaker indicates the outcome of the game to the learner.

An alignment is generated over the *perceived topic* and *intended topic* and the agents can update the names of the objects in their inventory. The naming game is regarded as a *communicative success* [84] when the listener points to the correct object.

In this section the principles denoting a successful communication in human conversation have been presented. These principles including a task focused communication and co-operation between the participants, which are also important elements in multi-agent systems and dialogue games.

In dialogue games communication can be modelled in terms of *what* is being exchanged within a message (i.e. the message's content) and *how* these messages are

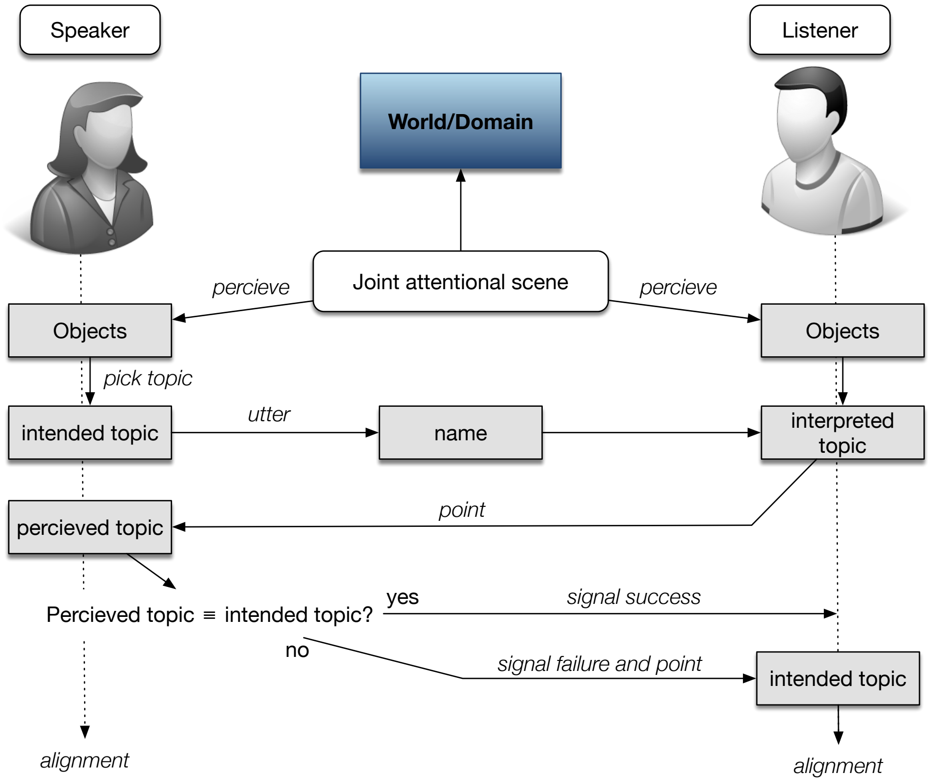


FIGURE 4.2: Agent communication example showing Naming Game [84].

exchanged between participants using locutions. These locutions can be modelled using *illocutionary speech acts* [10]. Finally the *modalities* present the rules of this dialogue which defines what is assumed of the participants involved.

As artificial intelligence developed, many of the ideas relating to human communication were borrowed and adapted from the fields of philosophy and psychology. Particularly the work characterising the notion of agents (as entities with decision making abilities) influencing the definition of software agents such as communicative acts described below.

In terms of a spoken conversation the same structure still holds as an ordered exchange between two autonomous agents, (i.e. agents capable of autonomous decision making) requiring at least two participants who strive to exchange information in order to achieve a common goal.

Speech acts present ways in which words can be used to present information and also actions and functions relating to the words themselves. Austin introduces the notion of three categories of speech acts, which are outlined as follows:

1. *Locutionary Act* is the act of actually 'saying something' [10], i.e. making a meaningful utterance, for example: *'Go Away!'*

2. *Illocutionary Act* is a complete speech act conveying a function, which denotes the act performed in the utterance such as: *'He told me to 'Go Away'.'*

3. *Perlocutionary Act* is the consequence of an utterance, such as: *'He got me to go away'*.

Austin uses *performative verbs* within his defined speech acts providing a specification in the action performed as a result of the utterance, these include *request*, *inform*, and *promise*. As discussed in Woolridge et al, Austin identifies a set of rules in order for a performative to be successfully completed [135]. These rules are referred to as *felicity conditions* [10](presented in [A.i..iii]) and can be related to Grice's co-operation principle and Pask's definition of strict conversation, in that these conditions must be met, in order for it to be successful. These *felicity conditions* include:

A.i There must exist an accepted conventional procedure specifically including the persons and circumstances involved;

A.ii The act must be executed by all participants both correctly and completely;

A.iii Acts must be performed truthfully, and completed by participants.

Searle extended Austin's work on speech acts and identified principles which must be upheld in order for a conversation between participants to be successful [115], presented in [S.i..v]. Searle discussed these principles in terms of a *speaker*, *listener*, and an *action*, and developed the following classification based on Austin's illocutionary speech acts:

S.i *Assertives*: commit the speaker to something being a fact, for example: *'No one runs faster than me'*.

S.ii *Directives*: try to make the listener perform an action, for example: *'Please can you open the door?'*

S.iii *Commisives*: commit the speaker to an action in the future, for example: *'I fly to Bologna tomorrow'*.

S.iv *Expressives*: express how the speaker feels about a state in context, for example: *'I am excited to see you'*.

S.v *Declarations:* change the state or condition of the world linking the speech act to the reality of the world, for example: *'I nominate you as candidate'*, with the consequence that the person is nominated as a candidate.

For messages to be understood by participants and for the dialogue between them to result in success, there are assumptions to be made regarding the agents which can be detailed as modalities. These modalities assign rules to the dialogue and truths that the participating agents can assume about each other.

For communication to be successful it must result in a contribution, where the participants involved co-operate with each other to achieve a mutual belief that the listener has understood the speaker well enough to continue or finalise the conversation. This notion allows the participants to be formalised in terms of roles within the given conversation to that of a *speaker* and a *listener*. These roles provide an ordering for the participants to make moves and co-operate with the other participants resulting in the achievement (or failure) of a goal. This co-operation between the two participants can be defined using Grice's co-operation principle [57].

Grice presented a co-operation principle allowing for a more formal representation of a conversation in terms of the moves available to a participant adhering to a set of given maxims in order for a conversational exchange to be successful.

Grice developed a set of four maxims ([G.i..iv]) which need to be adhered to in order to support his co-operation principle which is based upon human communication and the options available to a participant within a dialogue. These maxims are *quality*, *quantity*, *relation* and *manner*, all of which relate to the information to be exchanged between participants:

G.i The maxim of *quality* requires participants to be truthful regarding their responses within exchanges and that information is not put forward if there is no support. For example: in human communication a participant would only state the following if they believed this was true, and had corroborative evidence to support the statement:

'It's snowing.'

G.ii The *quantity* maxim refers to the participants exchanging the information as required that consists of enough to support a claim. For example if a waitress was to ask:

'How do you take your coffee?'

A response corroborating the quantity maxim could be:

'With milk and sugar.'

A response highlighting a break in this maxim would be:

'I like Arabica beans, grown in southwestern Ethiopia. I like a dark roast, served black with sugar and skimmed milk. I buy ground coffee beans and make espresso coffee when I am at home.'

This reply clearly shares more information than is required for the question asked. In contrast to this, too little information is a break in the maxim for example, replying to the above question with *'in a cup'*.

G.iii The *relation* maxim relates to the relevance of information that requires participants to only exchange information related to the current exchange. For example in human conversation: if one participant asked *'What is the weather forecast today?'* An inappropriate response would to this question would be discussing coffee beans.

G.iv The *manner* maxim refers to the ambiguity of information exchanged by participants. In keeping with the manner maxim if asked a 'yes' or 'no' question, a response would be exactly that rather than stating an ambiguous reply.

Pask in the field of psychology and education proposed the notion of *conversation theory* [104], a model of the underlying processes involved in complex human learning. The principle idea of the conversational theory states that learning occurs through conversations about a subject matter which allows knowledge to become explicit. In the model of conversational theory, Pask defines a *strict conversation* setting out a similar set of rules to Grice:

P.i The participants agree to obey the rules of the conversational language;

P.ii The conversation is focussed upon a conversational domain;

P.iii The conversational domain involves a particular type of representation of what may be known and what may be discussed;

P.iv Each topic is said to be learned within strict conversation;

P.v In this context, understanding is given a specific and technical connotation though the inputed meaning and corroborates with, and furthers the usual meaning;

P.vi Strict conversation is punctuated by understandings and the intervals occupied in reaching an understanding are called occasions.

These definitions ([P.1..vi]) and the similar maxims defined by Grice can be upheld as 'conversational rules' within information exchange in dialogues for autonomous agents. Take for example the interaction in Figure 4.1 between a teacher and a learner. A teacher draws the learner's attention to an object in an environment using the name 'ball' and the learner can deduce from the gestures and speech of the teacher, that the object in context is identified as a 'ball'. This however relies on the learners intrinsic ability to interpret the meaning correctly and link the meaning to the object. If this fails the learner is unaware of the context of the object and may incorrectly deduce that 'ball' is an attribute rather than the object itself. Similarity if the teacher was to just say 'ball' the learner would not have enough information to intrinsically perceive the meaning of the term resulting in this interaction failing. This is an example of a break in the quantity maxim defined above. It is important that within dialogical games these rules are formalised and followed by participants.

4.2 Dialogue games

A dialogue game is defined as *conventions of interactive goal pursuit* [86] where participants exchange utterances in accordance to the defined rules proposed in a protocol. These moves are exchanged by the agents until the dialogue terminates i.e. the goal such as generating an alignment has either been achieved or failed.

Addressing ontology alignment generation as a dialogue game is an approach which is gaining traction in the ontology matching and agents communities. Recent work such as [7, 18, 25, 40] have begun to utilise an iterative approach for example language games in order to address matching entities from one ontology to another.

The problem of communication based ontology matching was first introduced in [13, 14] where *ontology negotiation* was facilitated through a communication protocol that allowed agents to exchange ontological fragments by successively specifying the meaning of given entities. Other work has addressed different aspects of ontology negotiation [70, 81, 126]. The Anemone system by van Digglen et al advocates a minimal protocol whereby agents exchange logical definitions in an attempt to define a minimal shared ontology with no information loss [126]. However, it assumes that agents have perfect knowledge over the instances of their ontological models (*i.e.* the underlying approach was grounded through an extensional model), which was used to induce a class description covering certain instances.

Other approaches align heterogeneous ontologies through decentralised negotiation mechanisms [70, 105] or argumentation [78, 81]. In [70] agents selectively exchange details of a priori privately known correspondences, and propose repairs to address any emergent violations [69], resulting in alignments that are mutually acceptable to both agents without disclosing the full ontological model.

Using a dialogue game for generating ontology alignments provides agents with a decentralised approach allowing the agents to negotiate an alignment without disclosing their full knowledge bases to either a third party, or to each other. This incremental 'per needs' method of sharing, provides the agents with a degree of privacy where they do not have to share unrelated fragments of their ontologies, or fragments containing commercially sensitive knowledge to others. This allows the agents to negotiate alignments pertinent to a specific task, reducing the need to share unrelated knowledge to generate an alignment.

Dialogues can be modelled differently depending on the problems they are designed to address. Table 4.1 illustrates the six main dialogue categories detailed by Walton and Krabbe [131] showing the goals of the dialogue and the aims of the participants in co-operating towards this goal.

Using the dialogue categories from Walton and Krabbe [131], ontology alignment generation can be modelled borrowing from the definitions of both an inquiry and an information seeking dialogue. An inquiry dialogue as defined, is one which is initiated due to a general ignorance of a given *problem situation* [131] where the goal is to share

Dialogue Type	**Initial Situation**	**Main Goal**	**Participant aims**
Inquiry	General Ignorance	Growth of Knowledge and Agreement	Finding or disconfirming a 'Proof'
Information Seeking	Personal Ignorance	Spreading of Knowledge and Revealing Positions	Hide, Gain, pass or show personal knowledge
Negotiation	Conflict of interests	Making a deal	Getting the best outcome for 'oneself'
Persuasion	Conflicting points of view	Conflict resolution	Persuading the other participant
Deliberation	Need for action	Reach a decision	Influence outcome
Eristics	Conflict	reaching accommodation in a relationship	Win in the eyes of onlookers

Table 4.1: Dialogue categories related to ontology matching problem, taken from *Walton and Krabbe dialogue types* [131].

information in order to grow a knowledge base. In contrast an *inquiry* dialogue is similar to an *information seeking* dialogue however, is initiated due to personal ignorance, with the agents sharing and gaining personal knowledge in order to achieve the goal of spreading the knowledge they have.

In categorising ontology alignment generation into the form of an inquiry dialogue, the requirements of the problem as a dialogue game can be stated. As discussed in this chapter, for a dialogue to be successful it is a requirement that co-operation between two participants who are given specific roles must be adhered to. Co-operation specific for this task entails that the agents must assume their assigned roles, and share the appropriate information requested, in order to attempt to generate a meaningful alignment between their ontologies. Using ontology alignment generation as the inquiry dialogue the specific goal is defined as:

To generate a meaningful full or partial alignment between two ontologies, through the sharing of these ontologies by the participating agents.

The initial situation of the ontology alignment generation referred to as the *'OA'* dialogue game, is that both agents are characterised by a state of *'initial ignorance'* i.e. no knowledge of a prior alignment between the two ontologies exists to the agents.

4.2.1 General Dialogues

Conversation as an autonomous *AI* problem can be modelled in terms of a *dialogue game*, holding the key elements of conversation and speech acts theory described previously. These include the moves participants can make and the information exchanged between them. Hamblin [61] introduced the notion of a *formal dialectical system* which define the rules governing the moves and structures in an organised conversation, between at least two alternate turn taking participants.

A dialogue game defined as *conventions of interactive goal pursuit* [86] in which participating agents exchange utterances, known as *moves* or *locutions*, according to a set of rules known as a *dialogue protocol*. Each move has an unique name and can be used by the agents at a given state to contribute to furthering the dialogue towards a pre-defined goal. These moves are exchanged by the agents until the dialogue terminates and where the goal has either been achieved or failed. A dialogue is a series of arguments posed by participating players taking turns, where the first argument would be posed by the proponent then a second by the opponent until the goal of the game is completed.

Using this definition, dialogues can be used to address numerous agent problems and regardless of their application they can all be designed to include a number of main components which are illustrated in Figure 4.3:

1. A phase structure containing states which are traversed by the agents using the defined moves. Across all dialogues these phases will consist of at lease an *open*, an *information exchange* and a *close*.

2. A commitment store to log the moves made by the agents and commit the agents to the messages that are passed between them.

3. A task to which the agents are aiming to achieve as a goal.

4. A strategy available to the agents, allowing them to choose between moves available.

5. A protocol defining the rules dictating the moves available to the agents, and the roles they assume within the interaction.

6. If the dialogue requires argumentation or negotiation, there must be the notion of an argument, which can be agreed or refuted by the agents.

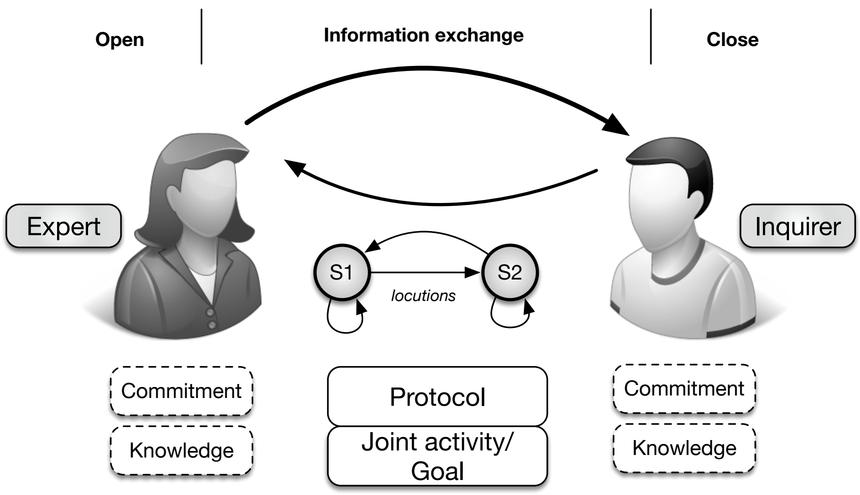


FIGURE 4.3: General Dialogue example

Dialogue Protocol and Agent Strategy

A dialogue game is defined as an exchange between two participants who utilise both a *dialogue protocol* and an *agent strategy* to outline the agents' decision making to successfully interact and communicate with each other.

A protocol presents a set of rules which regulate the interaction between the participating agents over the predefined goal. This protocol explicitly defines the moves designed as illocutionary speech acts that the agents are permitted to use throughout the dialogue. The dialogue protocol, as a fixed structure governs the 'turn' taking between the agents and assigns the agent roles, allowing them to successfully make moves and respond appropriately regardless of the agents strategic preferences.

The protocol is a public notion and is presented independent of any aims or goals the agents have which will define their choice of moves and must uphold the rule restrictions set within the protocol. The private notion of the protocol is assumed by an agent as a strategy presenting the reasoning used by the agents to asses the moves defined in the protocol.

The strategy is individual to the agents representing their personal aims and preferred outcomes from the dialogical interaction with the other agent. This strategy provides the agents with independent decision making *(DM)* methods which they utilise in order to chose one legal move over another throughout the dialogue.

Depending on the type of dialogue, the participating agents will adopt different DM strategies in order to get the optimum result concerning the individual agent's aims. Illustrated in Table 4.1 both of the participants aims are co-operative for example in an inquire or an information seeking dialogue. In contrast to this agent's aims can also be selfish as illustrated in the agent's aims for negotiation dialogue. If the agents adopt a co-operative aim the strategy will adhere to the *co-operative principle* such that the agents do not contradict the maxims defined by Grice.

The DM strategy the agents use will be influential in the outcome of the dialogue, as will determine what choices the agents make throughout this interaction and how the agents move between the various phases of the dialogue illustrated in Figure 4.3 and are presented as a series of pre-conditions defined in Chapter 5. DM strategies with conflicting agent objectives have previously been utilised within ontologies in the process of reuse, such as the MAUT approach [66] which identifies the best and most appropriate domains for reuse for a particular task in order to reduce costs which are associated with full and original ontology development [65]. DM approaches take a series of *attributes* as criteria in order to establish the importance or priorities within a each of the DM processes, in order to achieve the best result for the overall dialogue task.

Phase structure

Dialogues inherently present agents with moves to which they can utter in order to change the current state of a communication through locutions/moves defined in the protocol. These moves allow the dialogue to be represented as a state transition dialogue, which can be modelled over a set of phases traversed using these moves. As illustrated in Figure 4.3, regardless of the type of dialogue, a three phase structure consisting of an *open*, *information* and a *close* can be modelled. Using the generation of an ontology alignment as a dialogue goal, within this phase structure, two participants are given roles mirroring those defined by Searle: an *inquirer* and a *listener*. The agents enter into the dialogue at the *open* phase, whereby neither party has the full solution and have a set of locutions or moves, similar to illocutionary acts, defined in a protocol available to them. Throughout the *information exchange*, the agents adhere to a dialogue protocol defining the rules restricting the use of these locutions and travers the states sharing knowledge about their ontologies in the defined moves. The dialogue game is complete when the participants have actively engaged with each other and have arrived to a consensual solution, resulting in achievement or failure of a task. At this point the dialogue game enters a *close* phase.

Hulstijn further developed the three phases in order to structure a negotiation dialogue, modelled over the following five phases: *Open*, *Inform*, *Propose*, *Confirm* and *Close*.

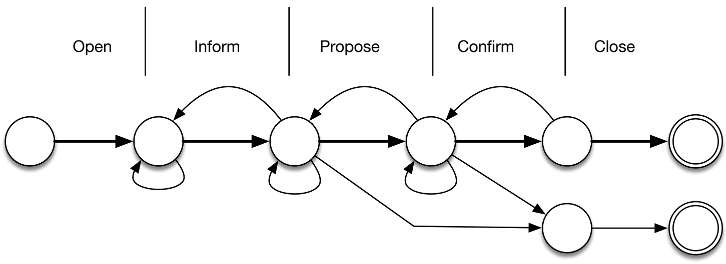


FIGURE 4.4: Hulstijn's dialogue phase model, used to develop the dialogical locutions for this work [64]

The five phases of dialogue shown in Figure 4.4, segment the transitions depicted by arrows which are the result of the joint action between the agents. The transition of the phases in Figure 4.4 corresponds to the exchange of several dialogue moves between the participating agents moving the dialogue process from one phase to another. No phase in the dialogue model can be skipped however they can be repeated. Within Hulstijn's dialogue the agents are assigned to three roles: an inquirer, an expert and a neutral observer.

The dialogue begins with an *open* phase which is the beginning of the exchange with the two participating agents agreeing to the rules of the dialogue and assign roles to the agents. It is in this phase that the negotiation space is established. It is in the *inform* phase of Hulstijn's dialogue model where there is 'mutual contact' with the participants and here they agree to the terms of the negotiation. In this phase the boundaries of the negotiation space defined in the previous phase are established. The *propose* phase of this five phase model provides the agents with the opportunity to reduce the negotiation space where the agents can accept of reject proposals. If a proposal is accepted, it generates a commitment between the agent and that proposal. The *confirm* phase sees the agents confirm or reject these proposals. This results in a positive outcome if the proposal is confirm or an unsuccessful outcome of the negotiation if the proposal is rejected. The *close* phase of the dialogue occurs regardless of the outcome established in the confirm phase. In this phase the participants close the current negotiation and re-establish themselves as interactive agents.

Commitment Store

Within a dialogue game the purpose of a *commitment store* is to document the utterances a participant makes and commits these utterances made by the the agents as true. Austin in detailing speech acts outlines the notion of inference in speech acts, and the concept of commitment similar to that defined by Walton [131].

Agents participating in dialogues, base their moves on what they know about themselves and what they assume about the other agents knowledge, based on the statements made by the other agent. This develops a separation between what is known to the agent and what an agent assumes about the other agent.

Hamblin introduces the importance of a store of *utterances* [61] between two participants, to separate an assumption and a fact. In order for a participant to negotiate over an uttered statement, a store of statements is required to be kept throughout a dialogue which provides a trace of what has been shared, and commits the agents to their utterances.

Walton and Krabbe further discuss the concept of a *commitment store*, in terms of propositional commitment defined as *'a kind of action commitment whose partial strategies assign dialogical actions that centre on one proposition'* [131]. Once posed, an agent becomes committed to an action in the context of the dialogue and as a result of a response, is therefore obligated to stand by the action and defend that commitment. For example if a speaker states to a listener *'a football is round'* and in response the listener states *'no, a football is oval'*, the speaker is then required to defend their initial statement with supporting statements.

Walton and Krabbe define rules of these commitments which mirror the previously detailed rules outlined in Section 4.1 of this chapter. These including the failure to comply with the rules of the dialogue, similar to the rule defined by Pask, *P.i* which could result in a loss of the exchange and failure to perform well within the dialogue, similar to the maxims in *G.i* to *G.iv*.

Formation of arguments

Unless a statement is made irresponsibly by an agent it requires an element of fact in which to support its validity. Depending on the type of argument, this support can be based on factual evidence providing a justification of the statement or a rebuttal providing an alternative statement to counteract the one initially posed. If there is a rebuttal this counter argument also requires support, to back the validity of this rebuttal over the initial statement made. In order to develop a structure to generate a meaningful argument Toulmin proposes a six part model, illustrated in Figure 4.5. This model includes the following parts: a *claim*, *data*, *warrant*, *backing*, *rebuttal* and *qualification*. The *claim* (C) represents the statement that is being made which will be supported by some element of fact. A *rebuttal* (R) is a counter arguments to these statements. If a claim is being made by an agent, it requires an element of factual evidence to denote the origin of the claim. The *data* (D) of a statement are the facts used in order to prove the argument where *warrants* (W) are used as logical statements linking the data to a claim. Using Toulmin's model, a warrant provides an explicit link in an argument between the claim and the data to which it relies on as its foundation. The *backing* (B) part of Toulmin's model, are statements used to support the truth of a warrants which can be expressed as a categorical statement of fact. In comparison to the data, the backing does not need to be explicitly produced within the argument. Finally *qualifiers* (Q) limit the strength of an argument by proposing explicitly the conditions under which the argument is true.

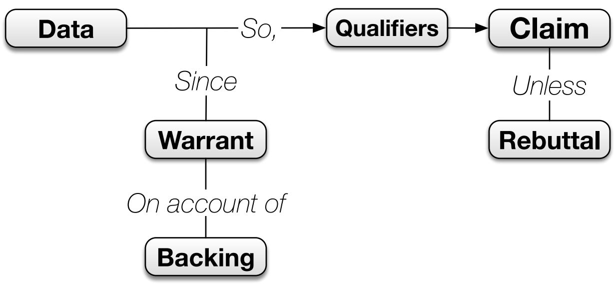


Figure 4.5: Toulmin's argumentation model [122]

Using the Toulmin model of argumentation, the following example illustrates the relations of the six part model [122]:

(D) Harry was born in Bermuda;
(Q) So, presumably;
(C) Harry is a British subject;
(W) Since a man born in Bermuda will generally be a British subject;
(B) On account of the following statues and other legal provisions;
(R) Both his parents were aliens/he has become a naturalised American.

Generating an ontology alignment through argumentation can be modelled using elements of Toulmin's argumentation model. If a candidate mapping (*a claim*) is proposed, the agents require knowledge (*data*) to establish semantic meaning (*warrant*) of the concepts in order to accept, reject or propose an alternative candidate mapping as a *rebuttal*.

4.3 Summary

This chapter has introduced the study of human communication with its origin in philosophy and psychology and presented an overview of communication from a cognitive perspective in human interaction.

The key principles of communication have been identified which are found in both the study of human communication and communication between autonomous agents. This chapter introduced the notion of a dialogue and what constitutes a dialogue from its origin in philosophy to its use in naming games [84] using multi-agent systems.

Communication as a dialogue game is formalised into a structured phase protocol providing the rules defining and the moves available where the participants interact over a given joint task (or goal). These moves are selected by the participants using a strategy, which allows independent decision making and dictate the traversal of the dialogue state by state. The notion of commitment has been introduced presenting the principle of a store of moves between the agents acting both as a log and committing the agents to the messages they have previously uttered.

The principles of communication between autonomous agents have been presented including Grice's co-operative maxims and Pask's rules of a strict conversation which provide assumptions about the participants. The way in which messages are shared between agents has been presented in terms of speech acts developed by Austin and Searle.

In Summary, this Chapter:

- Introduced the study of human communication and its origins.
- Identified the main principles that continue from human communication to that of multi-agent systems.
- Presented the roles of the participating agents within an interactive dialogue in attempting to address a pre-defined goal.
- Detailed the varying dialogue types as categorised by Walton and Krabbe and the phase structure to their design.
- Introduced the components of a dialogue including the *dialogue protocol*, *agent's strategy* and the notion of a *commitment store*, the formulation of an *argument*.

Part III

Contribution

Chapter 5

Dialogue Protocol

Chapter Outline

'We can only see a short distance ahead, but we can see plenty there that needs to be done'. - Alan Turing

Within a dialogue game participants must clearly understand the moves available to them in order to achieve a goal. The game is presented in a dialogical setting, where the joint goal is generating an alignment between the participating agent's assigned ontology, as illustrated in the previous chapter.

In this chapter the dialogue defining the interactions between the participating agents, for the purposes of ontology alignment is formally defined. This definition of the dialogue is represented over the properties of the dialogue, including a protocol defining the moves available to the agents in order to achieve the common goal of an alignment followed by a strategy defining the choice of moves the agents make at a given state in the dialogue.

5.1 DbMN Dialogue Components

Dialogue games introduced in Chapter 4 are used as a metaphor for a collaborative alignment mechanism where agents co-operatively form agreement. In this work dialogues provide a communication framework where agents can use a dialogical protocol rooted in human conversational processes, to investigate human based communication principles and rules and to utilise these principles for AI within a dialogue in addressing a task. This dialogue based approach was developed from an interest within the cognitive and psychological approaches to learning and human communication, and this approach was taken and adapted within an AI context to address the ontology alignment problem.

Figure 5.1 illustrates the conceptual architecture of the DbMN dialogue, and contextualises the components included in the approach. This figure presents how the two participating agents interact using the protocol, to address heterogeneity in ontology alignment generation.

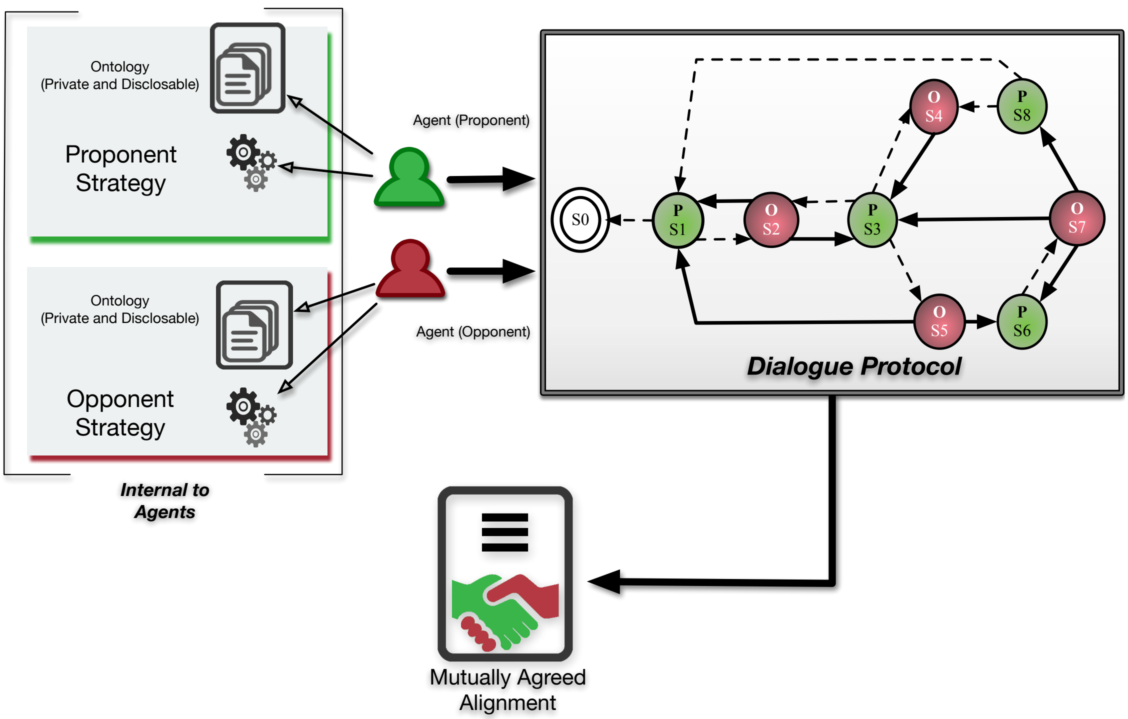


FIGURE 5.1: Conceptual Architecture illustrating how the use of ontologies, and the dialogue as a protocol and agent strategy fit into the approach, to address ontology alignment generation.

In this work a dialogue is defined in terms of moves and players [135] where the dialogue *moves* $\mathcal{M}$ is of type $\mathcal{T}$ movetype. These moves $\mathcal{M}$ are defined as a pair $\langle Player, Arg \rangle$. Here a *player* $\in \{0,1\}$ is specifically one of two agents making an argument within the dialogue and *arg* is the argument which is being made. The two agents and their fixed assigned roles are illustrated in Figure 5.1 as a *proponent* and *opponent* agent each with their respective ontology fragments and independent strategy which will dictate how they choose which move they will use at any given state within the dialogue. The agents also assume a role in each move determining which of the agents, in that given move is passing a message i.e. the *sender* and which agent is receiving the message i.e. the *receiver*. These roles of sender and receiver are interchangeable giving both the agents the opportunity to make a move, and thus co-operatively interact.

This chapter formalises the DbMN over the six dialogue components that are assumed within this approach and the decision making strategies used by the agents. These components have been introduced in Chapter 4 and include:

1. A commitment store used by the agents (Section 5.1.3);
2. An argument which is generated by the agents (Section 5.1.4);
3. A joint task over which the agents interact (Section 5.1.1);
4. A dialogue structure including phases and states over which the dialogue is modelled (Section 5.1.2);
5. A dialogue protocol (Section 5.2);
6. An agent strategy (Section 5.3.3).

5.1.1 Dialogue Task

In order to enter into the dialogue, the agent roles must be assigned to establish which agent is to move first. These agent roles of *Proponent* and *Opponent* are assigned at random at the beginning of the dialogue, prior to the utterance of the first locution. These roles are fixed, however the *sender* and *receiver* roles alternate between the agents throughout the dialogue, allowing both agents equal opportunity to share and receive knowledge pertinent to a candidate mapping.

This dialogue approach addresses the challenges arising in ontology alignment generation as a result of ontological heterogeneity and is illustrated in Figure 5.2. This illustrates the dialogue addressing the occurrences of entities modelled with conflicting meanings, which needs to be resolved in order to generate semantically meaningful mappings as seen within the *propose* phase detailed in the formation of agent arguments 5.1.4. Specifically the dialogue approach addresses semantic mismatches discussed in Section 3.3 including *Terminological* mismatches which are addressed by investigating the concept labels, and secondly *Content* mismatches which are established when exploring the concept relations. Similarly the protocol can be utilised if entity terms are represented different natural languages. Knowledge will still be shared between the agents, however a meaningful match might not be found unless there is a semantic similarity between the strings of the terms. In order to address this issue, an addition of an external resource such as BabelNet [96] could be included, however this is out of the scope of this thesis. The hypothesis and further assumptions made within this work are presented within this Chapter in Section 5.6 and Chapter 8

In review of the definitions presented in Chapter 2, ontologies are modelled as machine readable knowledge bases in the form of RDF [82] graphs, which are assumed by an agent. These ontologies are defined as *a formal, explicit specification of a (partially)* *shared conceptualisation.* In this context a *conceptualisation* refers to an abstract model of a world defined by relevant concepts and relations. *Explicit* means that the concepts and restrictions detailing the vocabulary are formally defined whilst *formal* is machine readable. *Partially-shared* denotes that the ontology assigned to an agent might not be shared in full. Privacy in ontology alignment generation has been previously investigated in [18] and the DbMN approach examines the concept of privacy by attempting to generate an ontology alignment where the ontologies are not fully disclosed.

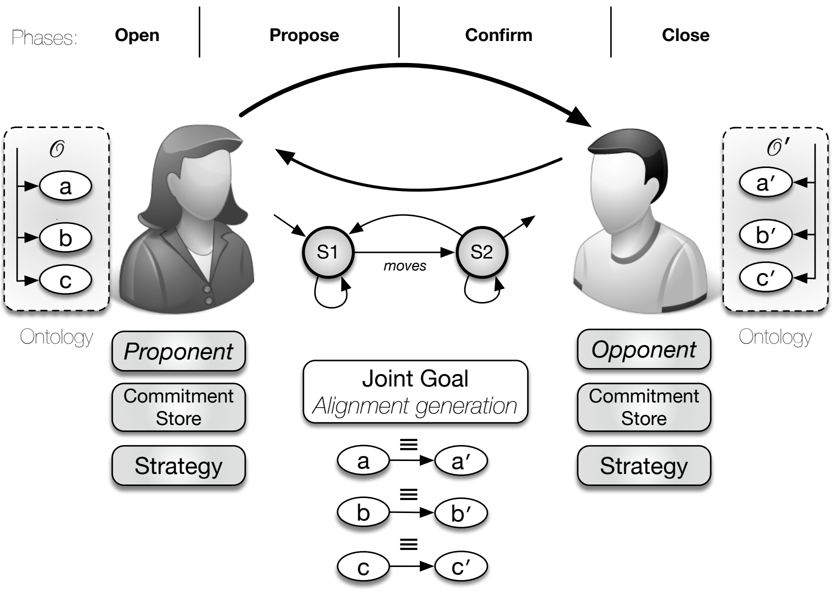


FIGURE 5.2: *DbMN* approach presented as an inquiry dialogical game

It is often the case however that agents differ in the vocabularies (ontologies) they assume, thus compromising seamless *semantic interoperability* between these dynamic and evolving systems.

This dialogue is described as a protocol which defines and formalises the moves available to the agents to address heterogeneity of concept names. This dialogue allows the participants to incrementally share the disclosable parts of their vocabularies Σ_d^x (*sender*) and $\Sigma_d^{\hat{x}}$ (*reciever*) which they are willing to reveal to the other agent, in order to mutually agree on a candidate mapping to form an alignment. These ontologies are modelled as a set of axioms describing classes and their relations, where $\Sigma = \mathsf{N_C} \sqcup \mathsf{N_R}$ is the *ontology signature*; i.e. the set of class and property names used in O' (the ontology definition is restricted to classes and roles, denoted respectively by $\mathsf{N_C}$ and $\mathsf{N_R}$).

The alignment method proposed assumes that the ontologies are represented as an edge-labeled directed graph G where G is an ordered pair $G = (V, E)$ such that:

- $V \subseteq \mathsf{N_C} \cup L$ is a finite set of vertices (where L is the set of literals);
- $E \subseteq V \times \mathsf{N_R} \times V$ is a ternary relation describing the edges (including label). As the direction of the edge $e \in E$ represents the 'subsumes' relation ($\sqsubseteq$), two edges are required to represent 'disjoint' ($\perp$) and 'equivalent' ($\equiv$).

The agents participate in this dialogue to establish a set correspondences (mappings) between entities in their disclosable ontologies Σ_d^x and $\Sigma_d^{\hat{x}}$. This correspondence is defined as:

Definition 3: *A* **correspondence** *is a triple denoted* $\varpi = \langle e, e', r \rangle$ *such that* $e \in \Sigma_d^x$, $e' \in \Sigma_d^{\hat{x}}$, $r \in \{\equiv, \sqsubseteq, \sqsupseteq, \perp\}$.

An ontology alignment defined in Chapter 3 determines the relationships between entities in two ontologies. However, in similar domains the ontologies can be modelled differently using a variety of modelling languages and contrasting assumptions making translating one ontology into another increasingly difficult i.e. resulting in semantic and syntactic heterogeneity as described in Chapter 3. It is stated in Chapter 4 that a dialogue requires the participating agents to communicate over an explicit task. The assumed task in the development of this dialogue is ontology alignment generation, and is defined in terms of the following main research question:

Research Question: *Can a "plausible" alignment between two ontologies be established, without sharing the ontology in full, prior to the matching process?*

In this work a "plausible" alignment is an alignment which contains mappings that a human domain expert would consider correct. Current ontology alignment systems participating in the OAEI, including those discussed in Chapter 3 evaluate the generated alignment, found by the system, to a reference alignment designed by domain experts. The mappings in a reference alignment however, may not be precise with respect to the constraints defined by a given task. A successful conversation in terms of the explicit task in this research question, would result in the agents generating a plausible alignment between their two assumed ontologies, which can be compared to a reference alignment to evaluate the accuracy and correctness of the included mappings, such that they are found to be semantically meaningful. Due to the heterogeneous nature of ontologies this task is not always straightforward. Finding a "correct" mapping is complex [24], thus evaluating alignments against a reference generated by domain experts is not always meaningful. In order for the dialogue to successfully complete this task, it must assume the following principles borrowing from the rules of a successful human conversation defined in Chapter 4. Thus the dialogue obeys Grice's *co-operative principle* [57] by assuming that:

i The participating agents are truthful;

ii They make informative contributions as required;

iii They keep their interactions terse and do not provide more information than necessary.

Furthering Grice's *co-operative principle* the dialogue also adheres to Pask's rules of a strict conversation such that:

i Participants agree to the rules of the dialogue in order to participate;

ii The dialogue is focused resulting in the success of a given task.

Making these assumptions and assigning these rules within a dialogue protocol, it is assumed the agents will be truthful in their statements and their decision making strategies will be co-operative, in order to successfully achieve the task of generating an alignment between their respective ontologies.

5.1.2 Dialogue Structure

The dialogical structure is a key component from the conceptual architecture of the approach, presented in Figure 5.1 and is based upon the phases designed in Hulstijn's model [64] for agent negotiation. Hulstijn defines negotiation as a *'constant process of raising and resolving issues'*, which can be used to structure the problem of establishing semantic meaning in the task of generating an alignment.

In contrast to Hulstijn's model the 'inform' phase is where the agreement of terms is established by the agents. In the dialogue presented here, this agreement is found in the initial 'open' phase allowing a four phase dialogue structure to be designed, excluding the inform. This dialogue structure is summarised in Table 5.1 illustrating the *speculative* and *confirmative* phases and the states in which the available moves can be utilised by the sending agents. This section will detail these phases and outline their design in terms of the agents and the dialogue task.

	Move	**Phase**	**State**	**Sender agent**
Speculative	*Initiate*	Open	S1	Proponent
	Propose	Propose	S2	Opponent
Confirmative	*Reject*	Confirm	S3, S7	Proponent or Opponent
	Justify	Confirm	S3, S5, S7, S8	Proponent or Opponent
	Testify	Confirm	S4, S6	Proponent or Opponent
	Assert	Confirm	S3, S7	Proponent or Opponent
	Accept	Confirm	S5, S8	Proponent or Opponent
	Fail	Confirm	S2	Opponent
	End	Close	S1	Proponent

Table 5.1: Summary of the structural design of the DbMN dialogue

Open: This is the beginning of the dialogue, and the start of the *speculative* part of the dialogue. Within the *open* phase the initial concept label is presented as the task to be mapped, by one agent to another. This initial move sets the first roles of 'sender' and 'receiver' and initiates a negotiation space. In comparison to Hulstijn's inquiry dialogue there are only two roles available for the agents, and there is no neutral observer. In contrast to the assignment of the roles of *proponent* and *opponent* which are fixed, the roles of a sender and a receiver switches throughout the phases in the dialogue, allowing both agents the opportunity to both send and request information.

Propose: In order for the agent to assert an argument about a candidate mapping, the grounds for that assert need to be developed by the agents and agreed upon in terms of an argument which comprises of a claim and support as defined in Chapter 4. This required support is defined and returned in the form of triples where the proposal is a potential string match of a concept label in $O1$ to a concept label in $O2$ and the subject of the triple is that concept label. This is the beginning of the development of a premise designed using the Toulmin model for argumentation, which is found in both the *propose* and *confirm* phases.

Confirm: The *confirm* phase of this dialogue, is the first of the *confirmative* phases and allows the agents to develop the claim of an argument by sharing elements of their knowledge bases as factual support backing up this claim. This occurs for the proponent and opponent agents allowing both agents to garner information for each other in order to develop support for the claim of the argument to be mutually accepted. Once the agents have the required support, the argument can be accepted in this phase of the dialogue.

Close: The *close* phase of this dialogue occurs when the agents either succeed or fail in finding an alignment between their ontologies. At this close phase, the commitment stores are cleared for both agents, allowing for a new dialogue to begin with new agents, or with the agents taking up the alternate roles of *proponent* and *opponent*.

5.1.3 Commitment Store

An important component of a dialogue game is the notion of a commitment store logging the utterances a participant makes. Within this dialogue there are two types of stores where the agent utterances are kept. These include a public stored shared by the agents (CS) and a private store (*Gamma store*) kept by each agent individually and is differentiated between the sender's store (Γ^x) and the recipient's store ($\Gamma^{\hat{x}}$). The agents manage these commitment stores which commit both agents to statements of knowledge sharing that has been sent from agent to agent throughout the dialogue process.

Within each of the gamma stores, the knowledge garnered is represented as a (partially) connected graph. This is the *neighbourhood* of the concept v_i, *i.e.* the subgraph originating from the vertex v_i constructed through the exchange of triples (ϖ) that form a directed path from v_i to support its candidacy, [$\langle v_i$, predicate, object$\rangle$].

The content of the private gamma stores Γ are updated as a result of a message being sent to a receiver within the following moves, illustrated in Table 5.2:

Testify in the form of: $\langle x, \textsf{\textit{testify}}, e, e', \varpi \rangle$ or $\langle \hat{x}, \textsf{\textit{testify}}, e, e', \varpi \rangle$

Assert in the form of: $\langle x, \textsf{\textit{assert}}, e, e', A \rangle$ or $\langle \hat{x}, \textsf{\textit{assert}}, e, e', A \rangle$

Walton and Krabbe state that a *commitment* is not always permanent and can be retracted with evidence [131]. Within this work however, the possibility of retracting

	Phase	*Message*	Γ^{x}	$\Gamma^{\hat{x}}$	*CS*
Speculative	Open	$\langle x, \textsf{\textit{initiate}}, e, \text{nil}, \text{nil}\rangle$	-	-	-
	Propose	$\langle \hat{x}, \textsf{\textit{propose}}, e, e', \text{nil}\rangle$	-	-	-
Confirmative	Confirm	$\langle x, \textsf{\textit{justify}}, e, e', \text{nil}\rangle$	-	-	-
	Confirm	$\langle \hat{x}, \textsf{\textit{testify}}, e, e', \varpi\rangle$	$\langle e, e', \varpi\rangle$	-	-
	Confirm	$\langle x, \textsf{\textit{assert}}, e, e', A\rangle$	$\langle e, e', \varpi\rangle$	$\langle e, e', A\rangle$	-
	Confirm	$\langle \hat{x}, \textsf{\textit{accept}}, e, e', A\rangle$	$\langle e, e', A\rangle$	$\langle e, e', A\rangle$	$\langle e, e', A\rangle$
	Close	$\langle x, \textsf{\textit{end}}, \text{nil}, \text{nil}, \text{nil}\rangle$	$\langle e, e', A\rangle$	$\langle e, e', A\rangle$	$\langle e, e', A\rangle$
		-	-	-	-

TABLE 5.2: Showing the update of agent's private and public knowledge stores throughout a dialogue run.

a fact from a commitment is not available to an agent and therefore commitment in the dialogue is permanent whilst the dialogue is open. Once a message is passed to a receiver, that message is stored within the receiver's gamma store ($\Gamma^{\hat{x}}$), until the dialogue is terminated. It is only when the *end* move is uttered at state $S1$ that the gamma stores for both agents are cleared.

Both agents manage a public knowledge base or *commitment store* (CS) alongside the private gamma store. Although the agents maintain individual copies of the CS these will always be identical and therefore there is no distinction between them. The CS contains a trace of all the moves uttered by each agent and is based upon the notion of an agent's commitment in the dialogue [131]. The content of the private CS stores can be seen in Table 5.2 and are updated as a result of a message being sent to a receiver, in the following move:

Accept in the form of: $\langle x, \textsf{\textit{accept}}, e, e', A\rangle$ or $\langle \hat{x}, \textsf{\textit{accept}}, e, e', A\rangle$

The CS, like the gamma stores is only cleared when the *end* is uttered at state S_1 indicating a termination of the dialogue. Walton and Krabbe also discuss the concept of a *commitment store* in terms of *propositional commitment* [131], where it is stated that once a statement is passed, an agent becomes committed to an action in context and as a result of a response, is therefore obliged to defend that commitment. This notion of *propositional commitment* is found in Γ in this dialogue, however in the CS, the information has already been supported and agreed upon by both agents, therefore neither are obliged to defend the content.

5.1.4 Dialogue arguments

The final component in the dialogue is based on the notion of an *argument* permitting the sender agent to propose a candidate mapping to the recipient, together with any supporting evidence. The recipient agent can in turn accept this argument, or request more support for this argument using the *justify* move. The dialogue mechanism utilises arguments which allow the agents to propose candidate correspondences (between the entities in their respective ontologies) and to justify them or refute them on the grounds of some evidential fact, provided as a means of support.

This protocol assumes the agents co-operate with each other in order to reach an agreement on the best correspondence mapping, thus satisfying the co-operation maxims stated by Grice and the rules defining conversational exchange. As a result of this assumed co-operation, agents can only make arguments that assert the validity of a new correspondence that was not previously disclosed, or question its correctness by stating an alternative correspondence for one of the same entities. As each new argument either introduces a new correspondence or states a new premise for an existing one, there is no possibility of cycles in arguments and thus the agents will either reach an agreement or they will reject the proposal.

The arguments are defined over the language $\mathcal{L}$, where each agent can form arguments about a candidate correspondence c and entities e in the disclosable signature Σ_d^x of their ontology and the triples describing e. $\mathcal{L}$ is the set of formulae ℓ defined by:

$$\begin{aligned} \ell &::= \langle \text{entity} \rangle \mid \langle \text{correspondence} \rangle \mid \{\langle \text{triple} \rangle\} \\ \langle \text{entity} \rangle &::= \text{e} \\ \langle \text{correspondence} \rangle &::= \text{c} \\ \langle \text{triple} \rangle &::= \varpi \end{aligned}$$

Hence $\mathcal{L}$ will contain statements about Σ_d^x, $\Sigma_d^{\hat{x}}$ and the correspondences mapping entities from one signature into the other.

The dialogue presented in this work borrows from the notion of support in Toulmin's argumentation model [122] where if a *premise* or statement is made regarding a proposed mapping, it requires supporting evidence in order to defend it. Therefore an *argument* expresses the relationship between the *claim* proposed by the acting agent and the *support* backing this claim. This also satisfies Grice's maxim of *relation* whereby the information shared is pertinent to the proposed argument. Using Toulmin's argument model, an argument (A) can not be asserted without a premise (Pr) supporting a claim (Cl). The premise in the arguments take the form of sets of triples when the agent establishes the validity of a correspondence c, it states an *argument* in favour of c, by proposing evidence or *support* based on the triples exchanged until that moment, that are stored in its *Gamma* store, resulting in matched ϖ pairs (ϖ, ϖ'). Each ϖ disclosed by one agent will have some similarity to zero or more triples ϖ' disclosed by its opponent, as illustrated by the example in Figure 5.3 (top) between two example sets of triples supporting a correspondence between the entities *Paper* and *article*.

The structural similarity of the neighbourhood of two entities can be computed over the set $\mathcal{NS}$, i.e. the set of all the matching (ϖ, ϖ') *pairs* that form a bipartite graph in Figure 5.3. The support is based upon some *injective matching* between the bipartite graph representing the triples in an agent's own ontology, and those disclosed by the other agent as part of the dialogue such that no triple from one ontology is paired to more than one triple in the other ontology resulting in a *one-to-one* mapping between the sets of triples, thus resulting in matched pairs (ϖ, ϖ') (illustrated in the bottom of Figure 5.3). This dialogue is restricted to a one-way direction of argumentation, meaning

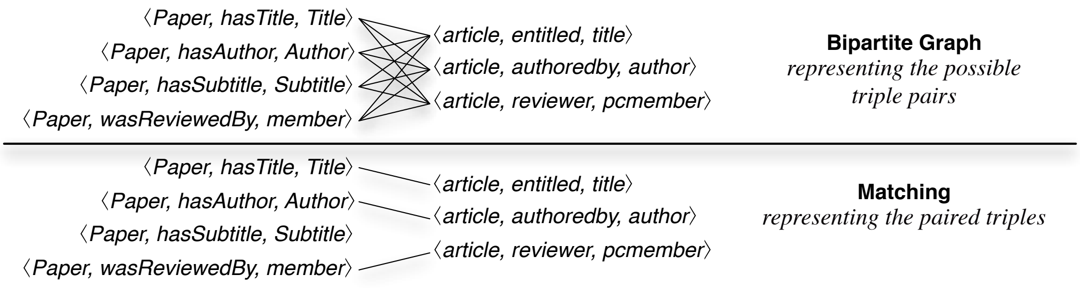


FIGURE 5.3: Possible pairs of triples (top) and a matching (bottom)

that potential mappings of a concept, once mapped are not revisited. Therefore a counter argument is not an option, the only locution available to an agent is to request support for a given argument thus no retraction occurs. Currently the assumption of the dialogue is that a mapping will only be found on the concept queried and support that comes from it in the form of triples. A further negotiation element to incorporate a rebuttal could be added into this dialogue as potential future work. The following triples found are illustrated in Figure 5.3 (bottom), once established they are not revisited. These include:

$$(\langle Paper, hasTitleTitle\rangle, \langle article, entitled, title\rangle)$$
$$(\langle Paper, hasAuthor, Author\rangle, \langle article, authoredby, author\rangle)$$
$$(\langle Paper, wasReviewedBy, Reviewer\rangle, \langle article, reviewer, pcreviewer\rangle)$$

The arguments within this dialogue can be formalised as:

Definition 4: *An* **Argument** *is a pair* $A = (Pr, Cl)$*, where* $Pr \subseteq \mathcal{L} \cup \{\top\}$ *and* $Cl \in \mathcal{L}$*. It is defined that* $\mathrm{Args}(\mathcal{L})$ *the set of all arguments derivable from the language* $\mathcal{L}$*.*

In this definition, Pr is the *support* (representing a set of premises of an argument) whilst Cl is the *claim*. Facts (i.e. statements with no premises) are represented as $(\top, Cl)$. An argument expresses a relationship between the *claim* and the *support* such that if the support holds then the claim must also hold. In the dialogue approach the support for an argument expresses a *justification* for some neighbourhood similarity (based on a set of related triples) for two entities e and e'. The claim asserts the viability of a correspondence between these two entities i.e. that the correspondence has some evidence of correctness.

5.2 Dialogue Protocol

The DbMN (*Dialogue based Meaning Negotiation*) protocol defines the rules of participation, which must be adhered to by the agents and has been previously published in [112–114].

The protocol depicted in Figure 5.4 was developed over the four phase structure detailed in Section 5.1.2 beginning with speculative phases followed by confirmation phases. The rules and locutions of the protocol within these phases were developed and refined and based on a number of principles from conversational theory including *strict conversation* rules defined by Pask and Grice's *co-operation principle*. The finalised DbMN protocol formalises the available moves and are designed based on the notion of *speech acts*, developed by Austin and Searle

Chapter 4 introduced dialogues as an approach to support one to one agent negotiation, where agents agree on the use of certain resources. Dialogue based negotiation assumes that the agents agree on a public 'language' for negotiation, whilst keeping their decision making strategies private. In this chapter the novel dialogue is proposed (DbMN), which allows agents to negotiate an alignment by proposing correspondences and their supporting evidence. Generating this ontology alignment using a dialogical approach requires a protocol to define the moves and a strategy to define the decision making process the agents will use to choose the moves. A dialogue game is defined as *conventions of interactive goal pursuit* [86] in which participating agents exchange utterances, known as 'moves' or 'locutions', according to a set of defined rules known as a 'dialogue protocol'. Each move has an unique name, and its contents is some statement which contribute to advancing the transitions in the dialogue towards a pre-defined goal. These moves are exchanged by the agents until the dialogue terminates and the goal has either been achieved or failed.

This section formalises the protocol in terms of the rules defining the moves available and followed by the decision making components which the agents use to traverse the dialogue. The DbMN approach makes a number of assumptions regarding the properties of the dialogue introduced in the previous chapter:

- There are two participating agents within the DbMN approach which both assume an ontology, and are assigned to the roles of proponent and opponent at the open phase of the dialogue;
- These ontologies are assumed to be machine readable;
- The agents are assumed to be truthful in their sharing of their knowledge bases throughout the dialogue, and their strategies must not contradict the rules set by the protocol.

The dialogue protocol comprises a sequence of communicative acts, or *moves* (denoted by $\mathcal{M}$ of type $\mathcal{T}$ movetype), whereby two participating agents take turns to share statements supporting or refuting a candidate correspondence. Figure 5.4 illustrates a state transition diagram of the dialogue showing the types of moves that the players can make and the choice of move which is then available in the new state.

The dialogue is designed over nine possible states where state $S1$ is the beginning of the dialogue and the *end state* is where the dialogue terminates. At state $S1$ the initiating agent known as the *proponent*, whose moves are represented with a dashed line, begins the dialogue with an *initiate* move. The set of legal *moves*, $\mathcal{T}$, are summarised in

Table 6.1 seen in Chapter 6 and are illustrated in the walkthrough example in Chapter 6. Figure 5.4 does not show the semantics of the moves, however the section below details these rules with respect to the pre and post conditions of each move available to an agent at a given state.

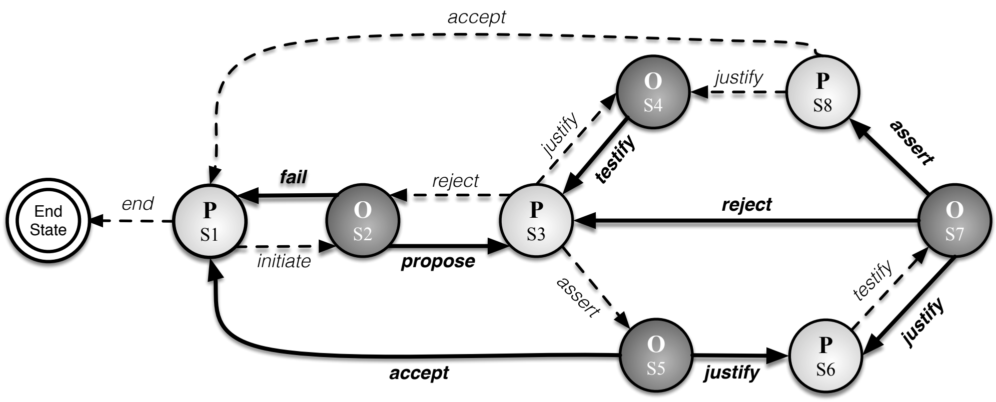


Figure 5.4: Dialogue Protocol detailing the flow of moves available to the Agents at each state.

For every dialogue move it is assumed that each agent plays a role; i.e. the proponent agent is either a *sender* x or *recipient* $\hat{x}$ (and conversely the opponent plays the alternate role such that they never play the same role concurrently). After each move the agents swap roles and thus take turns in acting as sender or recipient. At each point in the dialogue an agent selects from one or more moves depending on its *strategy*, which in turn is based on some *objective function* that reflects the agent's current task or objective. Thus an agent may want to find a maximal alignment (i.e. map as many entities as possible) if it is interested in knowledge integration or find some alignment that maps only those entities that are necessary and sufficient to perform some service [5]. When the proponent has no further entities to map, then the dialogue can terminate. If the opponent then wishes to explore further correspondences it can *initiate* a new dialogue and assume the role of *proponent* (i.e. the agents can reverse roles).

5.2.1 Dialogical Moves

This section discusses each of the moves available to the agents at each state of the dialogue as well as the pre-conditions and post-conditions of the move i.e. what is required for the move to be available to the agent and what moves can be used after the utterance is made.

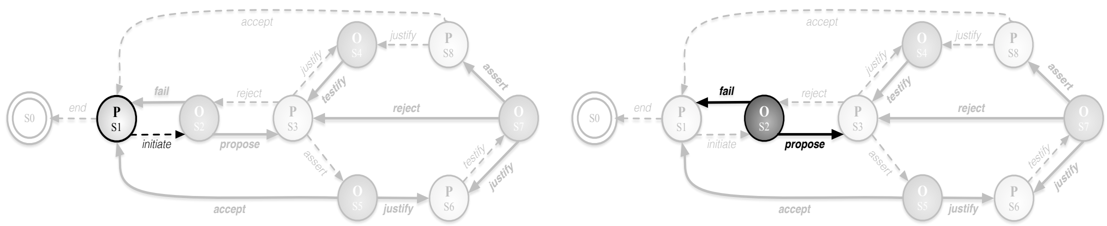
The syntax of each move is of the form $m = \langle x, \tau, e, e', l \rangle$, where τ is the move type such that $\tau \in \mathcal{T}$, and $\mathcal{T} = \{$*initiate*, *propose*, *assert*, *accept*, *reject*, *testify*, *justify*, *fail*, *end*$\}$; e represents the source entity being discussed (identified within the *initiate* move); e' is the current candidate target entity (i.e. the entity that could be mapped to e); and l represents a list of zero or more additional elements (depending on the type of move). For some moves it may not be necessary to specify the source entity the target entity or any additional elements, in which case they will be empty or unspecified (represented with *nil*).

Definition 5: *A* **dialogue** *denoted as $\mathcal{M}$, which comprises of a sequence of moves $\langle m_r, \ldots, m_t \rangle$, where $r, t \in \mathbb{N}, r < t$ are time points, involving two participants $x, \hat{x} \in \mathcal{P}$ where the roles of the participants are sender x and recipient $\hat{x}$, such that:*

- $\mathsf{sender}(m) = x \in \mathcal{P}$, *where* $m \in \mathcal{M}$
- $\mathsf{movetype}(m_r) = \textsf{\textit{initiate}}$
- $\mathsf{movetype}(m_t) \in \{\textsf{\textit{end}}\}$
- $\mathsf{sender}(m_s) \neq \mathsf{sender}(m_{s-1})$

As the dialogue progresses over time each move is denoted as m_s, $r < s \leq t$, where r is the time point of the first move of the dialogue, t is the time point of the last move, and s is the time point of the current move. Only participants in the dialogue can make moves, and the first move of a dialogue must always be an *initiate* move, and the last move should be an *end* move. Finally, each agent should take turns in uttering a move. Every move uttered is stored in both agents commitment stores such that $\mathcal{M} \subset CS$ where CS is the commitment store and therefore constitutes a fact that is know to both agents.

M.1 The move $m_r = \textsf{\textit{initiate}}$: $\langle x, \textsf{\textit{initiate}}, e, \text{nil}, \text{nil} \rangle$



(a) movetype $m_r = \textsf{\textit{initiate}}$

(b) $m_{r+1} \in \{\textsf{\textit{propose}}, \textsf{\textit{fail}}\}$

FIGURE 5.5: Initiate move, and legal post-conditions

An **initiate** move is uttered when a correspondence is desired for e, which must be in x's ontology; and should not have appeared previously in an *initiate* move. The only permissible subsequent moves are *propose* if the opponent has a potential match for the entity e, or *fail* if no potential match for e can be found. For the *initiate*, the following conditions hold:

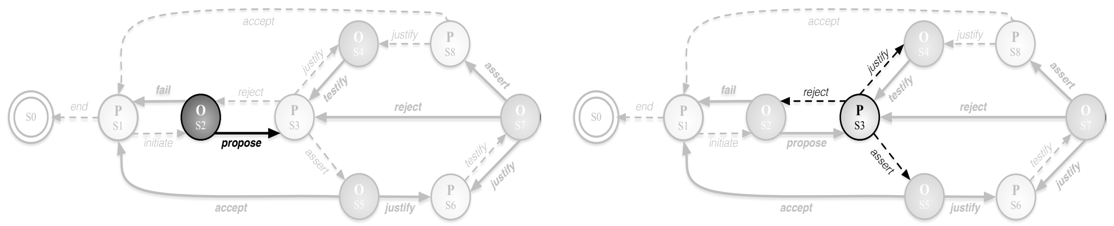
- Pre-conditions
  - $e \in O^x$
  - $\nexists A \in CS$ s.t. $e \in \mathsf{ent}(\mathsf{claim}(A))$
  - $\nexists m \in \mathcal{M}$ s.t. $(\mathsf{movetype}(m) = \textsf{\textit{fail}}) \wedge (e \in \mathsf{ent}(m))$
- Post-conditions

  * $\mathsf{movetype}(m_{r+1}) \in \{\textit{propose}, \textit{fail}\}$

A denotes an argument as a pair $A = (Pr, Cl)$ where Pr is the *support* (representing a set of premises of an argument) and Cl is the *claim.* Facts (i.e. statements with no premises) are represented as $(\top, Cl)$. This argument is defined formally in Section 5.1.4.

M.2 The move m_s = *propose*: $\langle x, \textit{propose}, e, e', \text{nil}\rangle$

The **propose** move is uttered when the sender x has some previously undisclosed target entity e' in its ontology which the sender has found is a *lexically viable* match for e such that the lexical similarity exceeds the lexical threshold value. (i.e. if $\sigma_l(e, e') \geq \epsilon_l$). The only permissible subsequent moves relate to the candidacy of some correspondence between e and e': *justify* is used if the opponent seeks further evidence to support the candidacy; *assert* if the opponent believes they have sufficient evidence in favour of the candidacy; or *reject* if no evidence can be found to support the candidacy. For the *propose*, the following pre and post conditions hold:



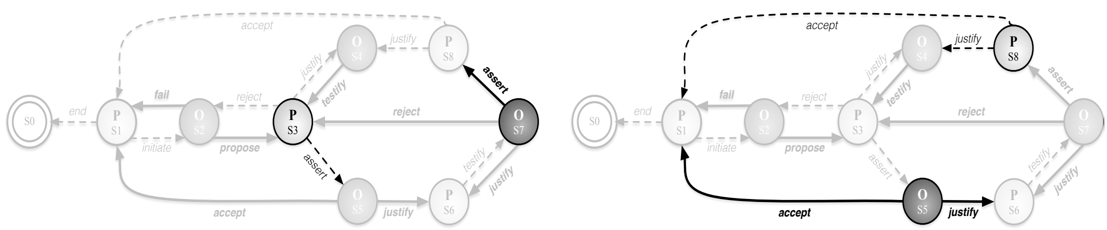
(a) m_s = *propose* (b) $m_{s+1} \in \{\textit{justify}, \textit{assert}, \textit{reject}\}$

FIGURE 5.6: Propose move, and legal post-conditions.

- Pre-conditions
  * $e' \in O^x$
  * $\nexists m \in \mathcal{M}$ s.t. $\{e, e'\} \subseteq \mathsf{ent}(m)$
  * $\sigma_l(e, e') \geq \epsilon_l$
- Post-conditions
  * $\mathsf{movetype}(m_{s+1}) \in \{\textit{justify}, \textit{assert}, \textit{reject}\}$

M.3 The move m_s = *assert*: $\langle x, \textit{assert}, e, e', A\rangle$

The **assert** move is made when the sender believes it has a candidate correspondence between e and e'; if i) the two entities are believed to be a *lexically viable* match; and ii) if the premise of the previously undisclosed argument A is considered *acceptable*; i.e. its aggregate *neighbourhood similarity* score σ_n (discussed in Section 5.3.1) is greater than or equal to the threshold ϵ_n. The only permissible subsequent moves are: *accept* if the opponent accepts the correspondence in the claim given its own assessment of the support or *justify* when the opponent rejects the support, and thus seeks additional or alternative evidence to support the claim. For the *assert*, the following conditions hold:



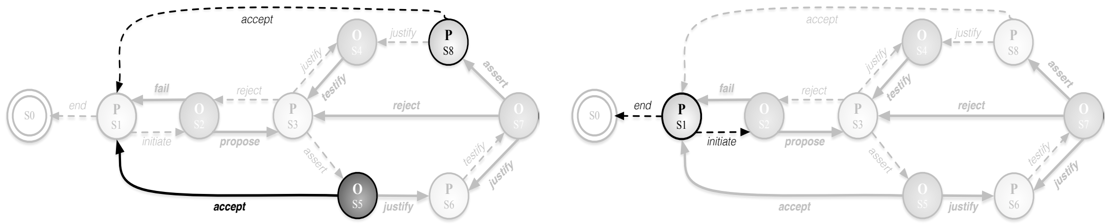
(a) m_s = *assert* (b) $m_{s+1} \in \{$*justify*, *accept*$\}$

FIGURE 5.7: Assert move, and legal post-conditions.

- Pre-conditions
  * $\sigma_l(e, e') \geq \epsilon_l$
  * $\exists Pr \subseteq \mathcal{NS} : \sigma_n(Pr) \geq \epsilon_n$
  * $\nexists m \in \mathcal{M}$ s.t.
    · $\mathsf{movetype}(m) = \textit{assert}$
    · $\{e, e'\} \subseteq \mathsf{ent}(m)$
    · $\mathsf{support}(m) \neq Pr$
  * $\mathsf{support}(A) = Pr$
  * $\mathsf{claim}(A) = \langle e, e', \equiv \rangle$
- Post-conditions
  * $\mathsf{movetype}(m_{s+1}) \in \{\textit{justify}, \textit{accept}\}$

M.4 The move m_s = *accept*: $\langle x, \textit{accept}, e, e', A \rangle$

The **accept** move is made when a candidate correspondence between the source



(a) m_s = *accept* (b) $m_{s+1} = \{$*initiate*, *end*$\}$

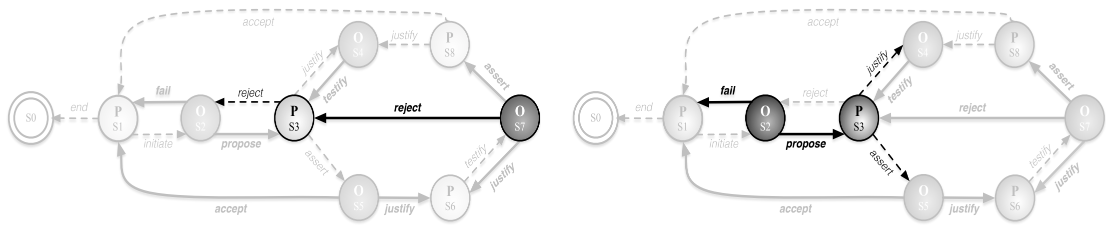
FIGURE 5.8: Accept move, and legal post-conditions.

entity e and the target entity e' is believed to be a *lexically viable* match and when the sender x believes there is sufficient evidence to support its candidacy. The argument A should be the same as that appearing in the previous move which should have been an *assert* move. Once this *accept* move has been made the argument A is then added to the commitment store CS. This is to prevent the entity e being subsequently proposed again in any subsequent *initiate* moves. The only move that is permissible in the dialogue following an *accept* move is an *initiate* move whereby a new source entity is considered. For the *accept*, the following conditions hold:

- Pre-conditions
  - $\sigma_l(e, e') \geq \epsilon_l$
  - $\exists Pr \subseteq \mathcal{NS} : \sigma_n(Pr) \geq \epsilon_n$
  - $\mathsf{movetype}(m_{s-1}) = \textit{assert}$
  - $A = \mathsf{contents}(m_{s-1})$
- Post-conditions
  - $CS = CS \cup A$
  - $\mathsf{movetype}(m_{s+1}) = \textit{initiate}$

M.5 The move $m_s = \textit{reject}$: $\langle x, \textit{reject}, e, e', \text{nil}\rangle$

The **reject** move is uttered when, x believes that there could be a *lexically viable*



(a) $\mathsf{movetype}\ m_s = \textit{reject}$ (b) $m_{s+1} \in \{\textit{justify}, \textit{assert}, \textit{fail}, \textit{propose}\}$

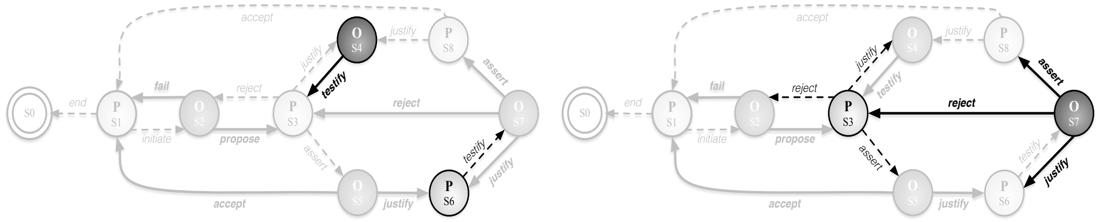
FIGURE 5.9: Reject move, and legal post-conditions.

match between the two entities e and e'; however, x does not have sufficient evidence to assert the candidacy of the correspondence; and it no longer believes the other agent has undisclosed knowledge that could support the candidacy for this correspondence. Thus there is no possibility of finding a correspondence for these two entities. The only permissible subsequent moves depend on the agent. If x is the opponent (S7) then the proponent can either *assert* an alternate argument for the correspondence or seek further evidence using the *justify* move. However, if x is the proponent at state (S3) then the opponent can utter *propose* if another entity could match e or *fail* if it is no longer possible to find a correspondence for e. For the *reject* the following conditions hold:

- Pre-conditions
  - $\sigma_l(e, e') \geq \epsilon_l$
  - $\nexists Pr \subseteq \mathcal{NS} : \sigma_n(Pr) \geq \epsilon_n$
  - Either one of the following must hold:
    - $\mathsf{movetype}(m_{s-1}) = \textit{testify} \wedge \mathsf{contents}(m) = \varnothing$
    - $\mathsf{movetype}(m_{s-1}) = \textit{reject}$
- Post-conditions
  - $\mathsf{movetype}(m_{s+1}) \in \{\textit{accept}, \textit{justify}, \textit{propose}, \textit{fail}\}$

M.6 The move $m_s = \textit{testify}$: $\langle x, \textit{testify}, e, e', \varpi \rangle$

The **testify** move is uttered in response to a *justify* move requesting evidence to



(a) movetype m_s = *testify* (b) $m_{s+1} \in \{\textit{justify}, \textit{assert}, \textit{reject}\}$

FIGURE 5.10: Testify move, and legal post-conditions.

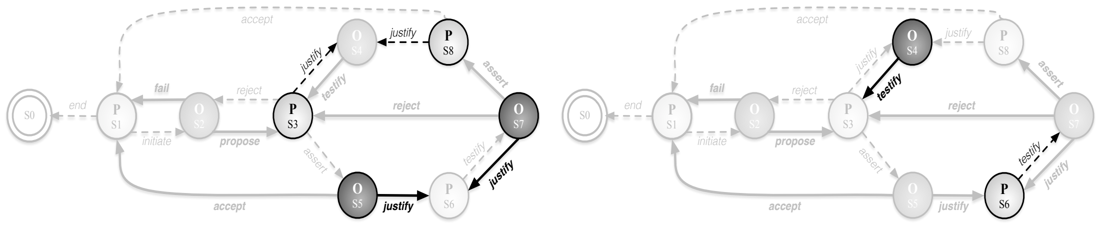
support the candidacy of some correspondence between e and e' which is believed to be a *lexically viable* match. The value of ϖ will be a new triple that supports the candidacy of the correspondence (if such an undisclosed triple exists) otherwise no triple (i.e. $\varnothing$) will be returned. The triple will be added to the recipient's commitment store $\Gamma^{\hat{x}}$. The only permissible subsequent moves relate to the candidacy of some correspondence between e and e': *reject* (i.e. no evidence to support the candidacy can be found); *justify* if the opponent seeks further evidence to support the candidacy; *assert* if the opponent believes they have sufficient evidence in favour of the candidacy; or *reject* if no evidence that can be found to support the candidacy.

For the *testify*, the following conditions hold:

- Pre-conditions
  - $\sigma_l(e, e') \geq \epsilon_l$
- Post-conditions
  - if $\exists \varpi' \in \mathsf{rank}^x(e), \nexists m \in \mathcal{M}$ where:
    - $\mathsf{contents}(m) = \varpi$
    - $\mathsf{movetype}(m) = \textit{testify}$

    then $\varpi = \varpi'$ else $\varpi = \varnothing$
  - $\Gamma^{\hat{x}} = \Gamma^{\hat{x}} \cup \varpi$
  - $\mathsf{movetype}(m_{s+1}) \in \{\textit{justify}, \textit{assert}, \textit{reject}\}$

M.7 The move $m_s = \textit{justify}$: $\langle x, \textit{justify}, e, e', \text{nil} \rangle$



(a) movetype m_s = *justify* (b) $m_{s+1} \in \{\textit{testify}\}$

FIGURE 5.11: Justify move, and legal post-conditions.

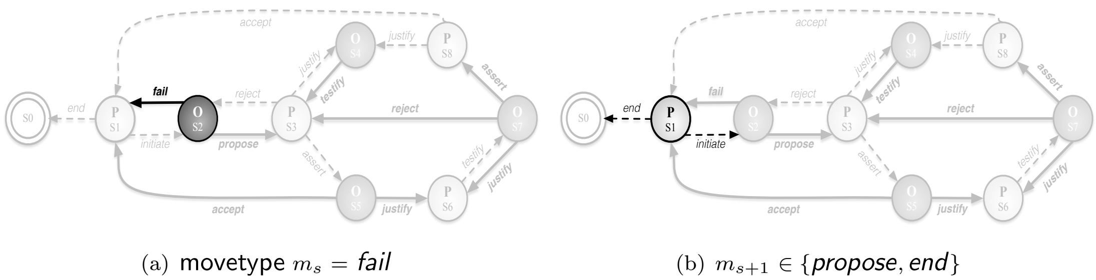
The **justify** move is uttered when the sender x believes the opponent S3 and S8 in Figure (a), or proponent S5 and S7 in Figure (b) may have disclosed knowledge that could support the candidacy of some correspondence between e and e'. Furthermore x also believes that there could be a *lexically viable* match between the two entities e and e' but does not have sufficient evidence to assert the candidacy of the correspondence. The only permissible move following a *justify* move is *testify*. For the *justify* the following conditions hold:

- Pre-conditions
  - $\nexists m \in \mathcal{M}$ s.t.
    - $\mathsf{movetype}(m) = \mathit{testify}$
    - $\{e, e'\} \subseteq \mathsf{ent}(m)$
    - $\mathsf{contents}(m) \neq \varnothing$
  - $\sigma_l(e, e') \geq \epsilon_l$
  - $\nexists Pr \subseteq \mathcal{NS} : \sigma_n(Pr) \geq \epsilon_n$
- Post-conditions
  - $\mathsf{movetype}(m_{s+1}) = \mathit{testify}$

It is important to note that the dialogue presented will always terminate when the agents enter into a *justify-testify* loop within the dialogue, due to their finite ontology signatures. If the dialogue does not end before every possible correspondence is considered from the proponent?s signature from the *initiate* move, then it will end, in its worst case, once all the knowledge to support a correspondence within the *justify-testify* moves has been disclosed. If no appropriate evidence is provided, then the dialogue will terminate following a *fail* move. If the agents had the possibility to retract an argument and pose a contradictory one using the alternative premise garnered between states S6 and S8, there would be a requirement for using a *justify* as the above pre-conditions would then hold. The current version of the dialogue has been designed with the possibility of adding this retraction. This retraction is not included in this version of the dialogue presented however, it is proposed as possible further work in Chapter 9.

M.8 The move $m_s = \mathit{fail}$: $\langle x, \mathit{fail}, e, \text{nil}, \text{nil}\rangle$

The **fail** move is uttered when the sender x is not able to find a target entity e' in



(a) $\mathsf{movetype}\ m_s = \mathit{fail}$ (b) $m_{s+1} \in \{\mathit{propose}, \mathit{end}\}$

FIGURE 5.12: Fail move, and legal post-conditions.

its ontology that has not been disclosed in a previous move and is a *lexically viable* match for the source entity e. This move signifies that no possible match for the source entity e could be found. The only move that is permissible following a *fail* move is an *initiate* move, whereby a new source entity is considered. For the *fail*, the following conditions hold:

- Pre-conditions
  - $\nexists e' \in O^x$ s.t.
    - $\nexists m \in \mathcal{M}$ s.t. $\{e, e'\} \subseteq \mathsf{ent}(m)$
    - $\sigma_l(e, e') \geq \epsilon_l$
- Post-conditions
  - $\mathsf{movetype}(m_{s+1}) = \textit{initiate}$

M.9 The move $m_t = \textit{end}$: $\langle x, \textit{end}, \text{nil}, \text{nil}, \text{nil} \rangle$

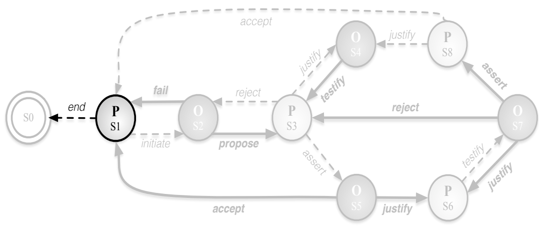


FIGURE 5.13: End move and legal post-conditions.

The **end** move is uttered in response to a *fail* or an *accept* move and ends the dialogue between the participating agents either with a successful alignment generated, if the m_{s-1} is an *accept*, or a failed mapping attempt meaning there is not enough support to corroborate a candidate mapping if the m_{s-1} is a *fail*. For the *end*, the following conditions hold:

- Pre-conditions
  - $\Sigma^t = \varnothing$
- Post-conditions
  - $CS = \varnothing$
  - Dialogue with current agents is terminated.

5.3 Co-ordinating dialogue moves through strategic decision making

Throughout the dialogue the agents require different decision making measures in order to accurately establish the similarity of entities of their own ontology, to those shared by the other agent in order to generate an alignment. These decision making measures include a lexical and structural similarity, a neighbourhood similarity and a ranking function defining which triple an agent will share. Finally the agent strategy which defines which move an agent will make at a decision point within the dialogue.

5.3.1 Dialogue Similarities

It is assumed that the ontologies in this approach, are represented in the form of a graph. The dialogue uses two similarity metrics: a *lexical similarity metric* and a *structural similarity metric*. These metrics are used in the different moves within the dialogue and handle two different types of knowledge. The lexical similarity measure $\sigma_l : \mathsf{N_C} \times \mathsf{N_C} \rightarrow [0, 1]$, uses a single concept label to establish both an *anchor* point and a path route through the graph. The structural similarity metric $\sigma_s : \Pi \times \Pi \rightarrow [0, 1]$, calculates the similarity of a triple in a 'receiver' ontology to those shared by the 'sender', in order to develop a premise Pr in an argument detailed earlier in this chapter.

Other measures used by the agents within the dialogue, include a *neighbourhood similarity metric* which compares the triples in an agents ontology O as a full bipartite graph to those shared by the other agent as a premise stored in the Γ to evaluate the strength of the support for an argument. The final measure used by the agents is an internal ranking mechanism defined as a *ranking function*, which is used in the dialogue in order for the agents to distinguish which part of their knowledge fragment they choose to share in a given move. This section will introduce these parameters abstractly and will be used in Chapter 7 more concretely through the use of an example illustrating the dialogue.

i **Lexical similarity metric:**

In this dialogical approach the agents can utilise different similarity methods such as the Jaccard similarity coefficient [98], or the Jaro-Winker [134] lexical similarity, or metrics that exploit linguistic resources such as WordNet [92] to identify synonyms, to determine lexical matchings. For the purposes of the evaluation of this approach a review of the comprehensive survey of different string similarity metrics [24] is presented in Chapter 7 to establish the choice of lexical matching method used to evaluate this approach.

For the purposes of defining this mechanism no assumption is made on the choice of similarity metrics used by the agents, nor is it specified that the agents have to agree on a common metric. Thus, it is assumed that agents differ in their assessment of the similarity of two labels.

Definition 6: *The* **lexical similarity metric** *is the function* $\sigma_l : \mathsf{N_C} \times \mathsf{N_C} \rightarrow [0, 1]$ *which returns the lexical similarity between the labels of two entity names* $e, e' \in \mathsf{N_C}$*, such that* $\sigma_l(e, e') = 1$ *iff* $e = e'$ *and 0 if the two labels are different.*

This function is used in the initial phase of the dialogue to discover those entities in opponent agent's signature that could lexically match an entity in proponent agent's ontology (*anchors*). A lexical match is considered *viable* if $\sigma_l(e, e')$ is greater or equal to its threshold ϵ_l. Thus, filtering out the cases where the similarity of e, e' is too low. The use of this lexical similarity between states S1 and S2 determines the root concept of the neighbourhood to be investigated within the ontologies.

ii **Structural similarity metric:**

This structural similarity metric uses *subject-predicate-object* triples ϖ that are shared by one agent, to allow the second to identify similar localised structures within their own ontology. This may be based on the triples in isolation or may also take into account other information that has so far been ascertained or inferred. As with the σ_l function, no assumptions about how the similarity function are defined at this point. Here this similarity function is assumed and for each agent and is defined formally as:

Definition 7: *The* **structural similarity metric** *is the function* $\sigma_s : \Pi \times \Pi \rightarrow [0, 1]$ *that returns the structural similarity between two triples* $\varpi, \varpi' \in \Pi$*, such that* $\sigma_s(\varpi, \varpi') = 1$ *if the two triples are considered as equivalent, and 0 otherwise.*

iii **Neighbourhood similarity:**

The DbMN approach requires a level of semantic coherence in order for a candidate mapping to be accepted into the alignment by the agents. This semantic coherence is established by developing a premise (Pr) by the agents sharing triples ϖ ($\langle subject, object, predicate \rangle$), which provide contextual evidence of the neighbourhood of a concept under negotiation. The agents can then compare the triples with their own knowledge base to establish this similarity of the neighbouring concepts.

The *neighbourhood similarity* σ_n is computed over the set of all matching (ϖ, ϖ') *pairs* (that form a bipartite graph seen in Figure 5.3), such that no triple from one ontology is 'paired' to more than one triple in the other ontology (i.e. finding an *injective*, or *one-to-one* mapping between the sets of triples). Depending on the choice of objective function used [49, 77], this can be achieved by finding a *matching* in the graph.

Definition 8: *The* **neighbourhood similarity** *is the function* $\sigma_n : \{(\varpi, \varpi') \in \Pi \times \Pi \mid \varpi \in \Gamma, \varpi' \in O\} \rightarrow [0, 1]$ *that returns an aggregate similarity calculated from a matching generated from the weighted bipartite graph obtained by calculating all possible structural similarities between the triples in an agent's gamma store* Γ *and the triples in the disclosable fragment of the opponent's ontology* $O^{\hat{x}}$*, such that* $\sigma_n(\varpi, \varpi') = 1$ *if the neighbourhood is structurally equivalent, and 0 otherwise.*

As no assumption is made regarding the objective function used to generate the matching (other than assuming that a structural similarity metric σ_s is used to generate the similarity of each pair), the function is defined as: $\mathsf{pairing} : \Pi \times O \rightarrow \Pi$ that generates a set of triple pairs given the triples in Γ and those in the agents ontology O.

For example taking the triples in Table 5.3 the proponent agent may have disclosed four triples to the opponent using an iteration of the *testify* move, within an exchange between the two agents. Therefore the opponent has the following triples learned from the proponent, along with their own knowledge base as illustrated in Table 5.3:

Gamma Store Γ^x	Ontology O^x
$\{\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title}\rangle$, $\langle \mathit{Paper}, \mathit{hasAuthor}, \mathit{Author}\rangle$, $\langle \mathit{Paper}, \mathit{hasSubtitle}, \mathit{Subtitle}\rangle$ $\langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member}\rangle\}$	$\{\langle \mathit{article}, \mathit{reviewer}, \mathit{pcmember}\rangle$, $\langle \mathit{article}, \mathit{entitled}, \mathit{title}\rangle$, $\langle \mathit{article}, \mathit{authoredby}, \mathit{author}\rangle\}$

TABLE 5.3: Triples sent throughout a dialogue run, stored in Γ^x and triples relating to a given concept in O^x

By using the structural similarity metric σ_s the complete set of all the possible triple pairs illustrated in can be determined. Assuming some objective function the matching can be generated. Thus, it is stated that:

$$\begin{aligned}\mathcal{NS} = \mathsf{pairing}(\Gamma^x, O^x) = \{ & (\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title}\rangle, \langle \mathit{article}, \mathit{entitled}, \mathit{title}\rangle), \\ & (\langle \mathit{Paper}, \mathit{hasAuthor}, \mathit{Author}\rangle, \langle \mathit{article}, \mathit{authoredby}, \mathit{author}\rangle), \\ & (\langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member}\rangle, \langle \mathit{article}, \mathit{reviewer}, \mathit{pcmember}\rangle)\}\end{aligned}$$

The premise Pr for the claim by the proponent agent for some correspondence c will comprise a subset of pairs from the set $\mathcal{NS}$ $\mathsf{pairing}(\Gamma^x, O^x)$, with a corresponding aggregate *neighbourhood similarity* σ_n. Although no assumption is made about how σ_n is defined it could be based on the structural similarity scores σ_s for each triple pair in Pr. A premise Pr is *acceptable* to an agent if $\sigma_n(Pr)$ is greater or equal to a threshold ϵ_n.

5.3.2 Rank function:

A key concept to this sharing within the DbMN approach is the notion that the agents may wish to maintain a level of privacy by not disclosing their ontology in full. Agents may also want to utilise a selective sharing approach if their ontology is very large, as it could contain entities relating to a broad domain, out of the scope of the joint task between the participants. Forcing an alignment with two unrelated concepts may result in a mismatch as discussed in Chapter 3. In contrast to traditional ontology matching methods using full *a priori* sharing of the participating ontologies, this work takes an incremental approach to ontology sharing, providing the agents with the option to not disclose their full ontology. The agents have both a *disclosable fragment* ($\Sigma^d \in O$), all of which they are willing to share on a per needs basis throughout the dialogue, and also a *private fragment* ($\Sigma^p \in O$) of their ontology, which they will not disclose in the interaction.

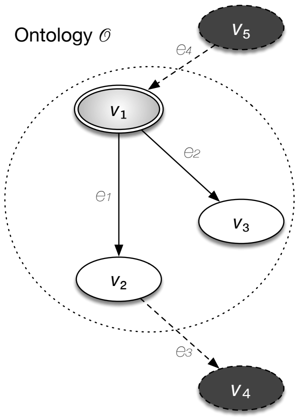


FIGURE 5.14: Ontology fragment presented as a graph, where nodes v_4 and v_5 are 'private' concepts.

Incremental Selective Sharing

Figure 5.14 illustrates the notion of the rank function for an ontology fragment O, which as a graph comprises a set of vertices V and edges e, such that $e \in E$ and $v \in V$. This figure depicts the disclosable Σ_d and private Σ_p fragments of the ontology O as $\{v_1..v_3\}, \{e_1, e_2\} \in \Sigma_d$ and $\{v_4, v_5\}, \{e_3, e_4\} \in \Sigma_p$. In the context of the ontology matching problem it is important that in the traversal of a neighbourhood of a concept v_1, the *subject of a triple* (ϖ) is going to provide the most semantically meaningful support in resolving the heterogeneity. The agents are attempting to maintain a level of privacy within this dialogue interaction, therefore this sharing method has been developed for the agents to share the fewest triples related to a concept, as support for a claim. This notion sees the agents ranking the similarity of the triples in the neighbourhood of the concept which will provide the most semantically meaningful support to back the claim posed in a premise Pr.

This graph traversal is how the incremental sharing between the agents is performed in the *justify* and *testify* moves in the dialogue using this *rank function* to share support for a candidate mapping. It is important to establish how the agents share structural elements of the ontology using the neighbourhood of an entity under consideration and it is here the *rank function* can be introduced.

For any given entity $e \in \Sigma$, there will be a directed path within the graph that relates e to other entities in its neighbourhood, where the maximum length of the path is bounded by the depth of the ontology. Given the example in Figure 5.15,

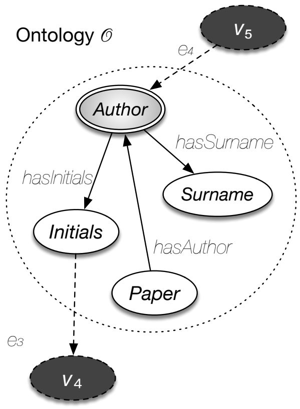


FIGURE 5.15: Ontology fragment presented as a directed graph.

the neighbourhood of *Author* would include the triple $\langle$*Author*, *hasInitials*, *Initials*$\rangle$, and $\langle$*Author*, *hasSurname*, *Surname*$\rangle$ but would *not* include the triple $\langle$*Paper*, *hasAuthor*, *Author*$\rangle$ as the concept *Author* is the *object* of the triple and not the *subject*. Thus, any triple (ϖ) within this path could be disclosed (i.e. shared with the other agent) to provide more details of the entities' local neighbourhood, provided that it forms a path from the entity itself. Depending on the strategy that an agent may adopt it may assume a depth-first traversal as opposed to breadth-first when disclosing its triples. Therefore, it is assumed that each agent utilises a function $\mathsf{rank}(e)$ that generates a strict pre-ordering of triples for a given subject e. This is formally defined as:

Definition 9: *The* **rank function** $\mathsf{N_C} \rightarrow \mathcal{R} \subseteq \Pi$ *returns an ordered list of triples in a path starting at some entity* $e \in \mathsf{N_C}$*, where* $\forall \varpi_i, \varpi_j \in \mathcal{R} : \varpi_i \not\succeq \varpi_j$.

An agent can request triples belonging to the local neighbourhood of some entity e' in the other agent's ontology to support the candidacy of a correspondence.

For the purposes of this definition no assumptions are made about how the ranking function is defined by any specific agent and thus the order in which the triples are ranked. Within the implementation of this dialogical approach, this rank function is specifically developed using four aggregated criteria.

These criteria include: *Subsumption*, *Rarity*, *Connectivity* and *Popularity*. These criteria allow the agents to rank the paths starting at a given entity, in terms of there lexical rarity and their 'centrality' in terms of a given concept. The criteria for developing the rank function is defined in the parameters in Chapter 7. This dialogue approach, follows the *Similarity Flooding* approach [91] the paths lengths are restricted to 1 and thus only disclose those triples for which e' is a subject.

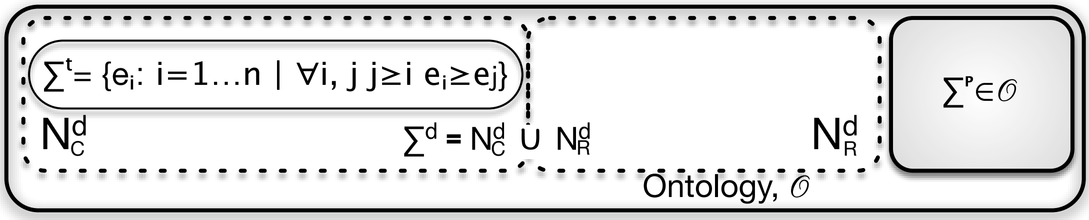


FIGURE 5.16: Dialogue Protocol depicting breakdown of disclosable agent Ontologies and signature.

5.3.3 Agent Strategy

As previously outlined in Section 5.2 the dialogue protocol allows the participating agents, a choice of legal moves over the states in order to traverse the dialogue. For example at *state* S1 the proponent agent can choose between the *initiate* or *end* locutions. An agent therefore requires a decision making strategy, allowing them to select one move over another from those available at any given state. These decisions are made at the states where there exists more than one possible action available to the agent, therefore the agent has to choose which best fits the overall argumentation strategy. This argumentation strategy is task oriented where the task is represented as a signature of concepts which need to be mapped.

This section borrows the formatting to describe semantic decision mechanisms used by McBurney et al, to detail the decision making mechanisms used by the agents at each decision point in the dialogue [73, 87]. The moves available to the agents at the corresponding states of the dialogue, are summarised in Table 5.4 showing at each state the agent performing the action and the available moves.

The task Σ^t illustrated in Figure 5.16 contains a set of concepts to be mapped. This signature is pre-defined by the *proponent* agent and requires the use of the dialogue to map *all* the concepts of this signature, in order for the task to be considered complete. This task is considered complete when a correspondence for each entity has been found which satisfy a level of confidence that is deemed acceptable by the agent, and therefore the proponent no longer needs the input from an opponent agent and can thus terminate the dialogue. The ordering of this task signature defines which entity is used in each iteration of an initiate move opening the dialogue. The agents both share the same task of aligning *all* concepts in a signature, with a *given level of confidence*, and with *minimal disclosure*. The proponent is defined by the agent initiating the dialogue over a Σ^t which is a subset of their ontology. The proponent agent in the dialogue has a disclosable set of entities in a signature $\Sigma^t = \{\mathsf{e_i} : i = 1..n | \forall i, j, j \geqslant i, e_i \geqslant \mathsf{e_j}\} \in \mathsf{N}_\mathsf{C}^d$ which, within the dialogue will be shared in the *initiate* move beginning with the first in the array of the signature. This initiate move will iterate through the ordered signature set to the next previously undisclosed entity, until all the entities in this ordered set have been mapped to any acceptable level of confidence. It is the opponent agent's job to comply with the

State	Agent	Available Moves		
S1	P	initiate	end	-
S2	O	propose	fail	-
S3	P	justify	reject	assert
S4	O	testify	-	-
S5	O	justify	accept	-
S6	P	testify	-	-
S7	O	justify	reject	assert
S8	P	justify	accept	-

TABLE 5.4: Available moves to aide decision mechanisms for proponent and opponent agents

dialogue rules and assist truthfully in finding a lexically similar concept to share in the *propose* move. The two agents, each possess a disclosable ontological fragment;

$$\Sigma^d = \{\mathsf{e_i} : i = 1..n\}_\mathsf{C} \cup \{\mathsf{r_i} : i = 1..n\}_\mathsf{R}$$

This signature is used by the agents in the dialogue fragment to find a correspondence c between two entities for each of the concepts in Σ^t.

Both agents implement the structural similarity metrics σ_s, defined in Section 5.3.1 which is used to compare similarity of the triples shared from the other agent within the dialogue at the *testify*, *justify*, *assert*, *accept* and *reject* moves or states 3-8. Note that these similarity pairs are not generated a priori, but are calculated during the dialogue. In order for a triple to be accepted, using the neighbourhood similarity metric, the σ_s value must meet a given threshold. The neighbourhood similarity metric $\sigma_n(Pr)$, defined in Section 5.3.1, calculates the average structural similarity σ_s of the triple pairs in the premise Pr.

5.3.4 Decision mechanisms for the proponent agent

The proponent agent's decision model is made up of four decision mechanisms over states S1-S8, seen in Table 5.4, with the proponent as the 'sender' agent at states *S1, S3, S6* and *S8*. These decision mechanisms allow the proponent to use the dialogue in order to negotiate over correspondences for the entities in their task, until all are mapped to a level that satisfies their argumentation strategy and complete a partial ontology alignment.

S1 **Begin negotiation, continue negotiation or terminate:** This mechanism allows the agent to begin, continue or end the dialogue over an initiated entity. Prior to this move the agent has a defined task signature Σ^t, all of which needs to be mapped. The concepts in Σ^t are randomly ordered prior to the opening of the dialogue and then used in this first move. If this task signature is not an empty set the dialogue can start with the proponent agent uttering an *initiate* move at state S1, stating a correspondence c needs to be found over a first given entity in the agent's task. If an *accept* has already been found over an entity in the Σ^t, the agent has two choices, to *initiate* if $\Sigma^t \neq \varnothing$ or *end* if $\Sigma^t = \varnothing$ (which is logged in the CS[1]). If this is at least the second *initiate* uttered by the proponent agent, meaning the dialogue has already been used to find a correspondence for an entity in Σ^t, the next ranked entity, which has not been previously shared by the agent will then be put forward in this move until Σ^t is an empty set, indicating that all the concepts have been mapped to a level the proponent agent is willing to accept. However, if no mapping can be found on an entity within the task signature, the task has failed as it specifically states *all* concepts in the task signature needs to be mapped. This will give the agent the opportunity to trigger a terminate in the dialogue, uttering the *end* move.

If the signature Σ^t of the proponent agent is an empty set, meaning a partial alignment of the signature concepts is complete and the task is achieved successfully, this will allow the agent to terminate the dialogue, (for implications of failure see state S2).

S3 **Evaluating or rejecting a correspondence and establishing a premise:**

This mechanism allows the agent to evaluate the correspondence proposed by the opponent. At state S3, the proponent agent has the choice between three possible legal moves regarding the confirmation of a proposed correspondence. These legal moves include: rejecting, accepting or requesting justification on this correspondence, all which use the following restrictions. The first time the proponent reaches state S3, the agent has a choice of *reject* or *justify*. At this point an *assert* is not a viable move as it requires a premise Pr which provides support of structural knowledge in the form of triples, which at this point has not been shared however, is garnered through the *justify*, *testify* loop between states S3 and S4. At this first traversal of state S3, if the lexical similarity for the proposed correspondence meets the proponent agent's ϵ_l, the *justify* move can be uttered in order for this agent to request neighbourhood support for the candidate mapping shared by the opponent in state S2 using the *propose* move. Alternatively the *reject* move can be uttered if the lexical similarity for the proposed correspondence is not met.

When the dialogue has previously reached state S3 (logged in the CS) the proponent agent has the choice between the *justify*, *reject* and *assert* moves, for a candidate mapping to be asserted in the dialogue, the argument requires support of a premise Pr. A *reject* move is uttered when the proponent can find no acceptable level of lexical similarity between the entity shared in state S2 by the opponent, as a candidate mapping to the entity posed in the *initiate* move.

If the lexical similarity for the proposed correspondence is above the proponent's lexical threshold (ϵ_l) it is then accepted as a potential candidate mapping however,

[1] If this is the first time the *initiate* move has been uttered the commitment store CS at this point will be empty. If there exists previous *initiate* moves, they will be logged in the CS and as a result of the formalisation each *initiate* will be on a new concept in the pre-defined order

it requires neighbourhood support which is when a *justify* move is uttered. If a *justify* move is being uttered it means that $Pr = \varnothing$ or $Pr < \sigma_s$.

In order for an *assert* to be made, the candidate correspondence needs to be supported with acceptable levels of neighbourhood similarity support. This is done using a premise Pr (in this dialogue, Pr is in terms of supporting triples). The structural similarity metric function σ_s measures the similarity of the triples anchored to the entity in question, at the *subject* of the triple which needs to meet a given threshold (ϵ_n) set by the individual agents.

If the previous move was a *testify* from state S4, the agent will take in the shared triple and use the structural similarity metric σ_s and asses the triple's similarity with those in its own ontology. The triple found with the highest similarity will be shared as part of the premise, in the *assert* from this state.

If the ϵ_n is not met the agent will require a further *justify* move to be uttered in order to garner this support. This *justify* will follow a *testify* move uttered by the opponent if $Pr < \epsilon_n$. When $Pr \geqslant \epsilon_n$ the proponent can utter an *assert*.

S6 **Supporting a developing premise:** This mechanism allows the agent to provide support for the opponent to develop an acceptable premise for the correspondence under negotiation. Here the agent searches through their own knowledge base to extract relevant triples to develop this premise.

At state S6, the proponent only has the option of a *testify* move. This *testify* move, operates on triples and will be used until there are no more triples left with the correct subject s. This subject s of the triple is the concept lexically similar to the concept being negotiated over from the *justify* move. The order in which the triples are shared, will be defined using the rank function over the set triples where the subject s of the triple is lexically equivalent to the concept under negotiation. When there are two or more triples each with the same ranking they will be selected and shared at random. If there are no results of a lexically similar match, the agent will return nil leaving the opponent with the decision to evaluate this response, seen in state S7.

S8 **Confirming, or continuing negotiations over a correspondence:** This mechanism allows the agent to confirm a correspondence and generate a partial alignment for the given entity under negotiation from the task signature or continue further negotiations.

Similar to the move choices at state S5 if an *assert* has been made by the opponent agent then the proponent agent has the choice between a *justify* and an *accept* move.

If there is enough supporting evidence of the candidate correspondence σ_s to meet ϵ_n, then an *accept* move uttered on the claim asserted at state S7 then a partial alignment can be generated including the correspondence and the premise updated in the CS. If there is not enough support to accept this correspondence, a *justify* is used so the proponent can garner further evidence in which to support this candidate correspondence under negotiation.

5.3.5 Decision mechanisms for the opponent agent

Alongside the proponent, the opponent agent's decision model is also made up of four decision mechanisms over states S2-7. These allow the proponent to use the dialogue in order to negotiate over correspondences for the entities in the proponent's task and assist truthfully until the proponent's task is considered by them to be complete. Only when this initial task is complete can the opponent agent initiate a new dialogue, assuming a task signature with entities to be aligned is available.

S2 **Considering a candidate entity:** This mechanism allows the agent to search through their knowledge base for a candidate entity they deem lexically viable as a match for the entity shared in the initiate move. This level of acceptability and the matching mechanism will be dependent on the agent's argumentation strategy.

At state S2 the opponent agent has the choice between two legal moves, the *propose* or *fail* moves. A *propose* is uttered when the opponent agent identifies that there exists at least one entity in their knowledge base that meets a lexical threshold level (ϵ_l) allowing it to be put forward as a candidate mapping. This lexical matching method is dependent on the agent's argumentation strategy and can be any lexical based method used in current ontology matching literature and does not have to be the same as that used by the proponent agent in the dialogue. If more than one entity is found the agent shares the highest candidate entity within it's own ontology which meets (ϵ_l) as a possible match for the entity shared in the *initiate* move. If a *reject* has been uttered traversing the dialogue from state S3 to S2, the opponent agent then shares the next highest candidate entity in its ontology deemed lexically similar, until there are no more that meet the (ϵ_l). If there are two or more which share the same lexical similarity then the entity shared will be selected at random.

If this threshold is not met i.e. the agents can not find an entity deemed lexically similar, this is when the opponent agent would utilise the *fail* move, indicating that there exists no entity in Σ^d the opponent considers acceptable that can be shared as a candidate entity mapping to the concept in the *initiate* move.

If the *fail* move is uttered by the opponent agent, the dialogue for mapping the Σ^t has failed as the strategy defines that all concepts in the signature must be successfully mapped. This move will lead to a termination of the dialogue.

S4 **Supporting a developing premise:** This mechanism allows the agent to provide support for the proponent to develop an acceptable premise for the correspondence under negotiation. Here the agent searches through their own knowledge base to extract relevant triples to be shared to develop this premise.

At state S4, the opponent agent only has the option of a *testify* move. This *testify* move operates on *subject-predicate-object* triples which are directed sub graphs of the agent's ontology.

This move will be uttered until there are no more triples left in the opponents ontology considered relevant. A triple is considered relevant when the subject s of the found triple, is lexically equivalent (this matching method is dependent on the agents argumentation strategy) to the entity under negotiation. The order in which the found triples are shared will be defined using the ranking method the opponent has used for the relevant triples. When there are two or more triples each with the same ranking they will be selected to be shared at random, but never shared more than once.

The implementation of this approach also utilises a version of the dialogue which shares all of these neighbourhood triples in a single testify move (a batch sharing approach), contrary to the decision mechanism here, presented as an incremental approach. The response options for the proponent agent remains the same, however the similarity is calculated over the set of all the given triples in the and the best is found in a single iteration of the justify-testify loop.

S5 **Confirming or continuing negotiations over a correspondence:** This mechanism allows the agent to accept a correspondence and generate a partial alignment for the given entity under negotiation from the task signature or continue further negotiations using the following restrictions. The first time state S5 is reached the agent only has the legal use of a *justify* move, as the opponent agent currently has no evidence of their own i.e. no shared triples from the proponent agent. The restrictions on the claim of an argument in this dialogue mean that any claim needs to be corroborated with *both* lexical *and* structural support therefore any *assert* at this time can not be accepted by the agent, leaving the agent with only one legal move. This *justify* move allows the agent to garner evidence which will allow them to accept or reject a candidate correspondence.

If the dialogue has previously traversed state S5, and an *assert* has been made by the proponent agent. The opponent agent has the choice between two legal moves, a *justify* or an *accept* move.

If there is enough supporting evidence for the candidate correspondence as an equivalent match in the premise of the argument asserted by the proponent i.e. the threshold ϵ_n is met (for ϵ_n to be met ϵ_l holds also), then an *accept* move can be made. If ϵ_n is not met, then a *justify* is used so the opponent can garner further evidence in which to support the candidate correspondence.

S7 **Evaluating or rejecting a correspondence and establishing a premise:** This mechanism allows the opposite agent as in state S3 to evaluate the correspondence proposed by the proponent, where the decision criteria is dependent on the agent's argumentation strategy. Here the agent can accept and establish or reject a correspondence or alternatively evaluate a premise using the following restrictions.

At state S7 this time it is the opponent agent who has the choice between the *justify*, *reject* and *assert* moves. A *reject* move is uttered when the opponent can find no acceptable level of structural similarity (i.e. $Pr < \epsilon_n$) between the triples shared in state S6 by the proponent as support for the candidate mapping.

Similar to state S3, for a candidate mapping to be asserted in the dialogue the argument requires support from both lexical and structural levels. Until these levels are met, no assert can be made leaving the agent with only the *justify* and *reject* as legal moves. If the agent has the required support for both the lexical and structural levels they can make an *assert* on the candidate correspondence.

If the ϵ_n is not met the agent will utter a further *justify* move in order to garner this support. When the structural similarity σ_s is indeed above the opponent's threshold (ϵ_n) it is then accepted as a potential candidate mapping and an *assert* can be made. If not a *justify* is repeated.

The use of lexical support alone could only be used when a *testify* triple is shared as nil meaning there exists no further triples relating to the entity, however some have already been shared e.g. at state S7, or at least a second iteration of *justify*, *testify* in state S3. This would be dependent on the agent's argumentation strategy.

These individual decision making mechanisms both for the proponent agent and the opponent agent alongside the dialogue protocol and specific locution rules detailing the legalities of moves at a given state, generate a framework allowing the agents to negotiate over a task specific. This task is to selectively generated incremental partial alignment. The subsequent Chapter 6, details an example walkthrough of the use of this dialogue, using two trivial ontology fragments between the proponent agent *Alice* and the opponent agent *Bob*.

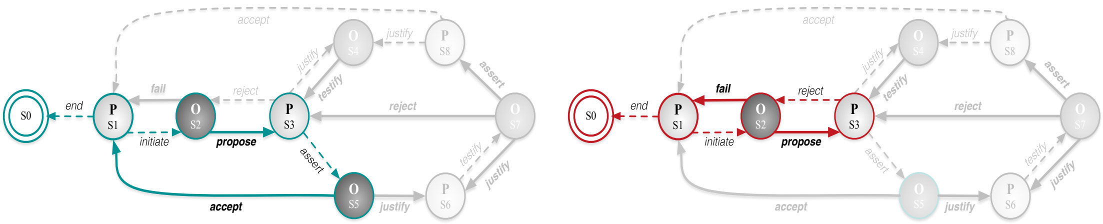
5.4 Dialogical Variants

The dialogue presented in this chapter, can be adjusted depending on the restrictions of the protocol regarding accepted mappings, agent decision making, and the proponent's signature. These variants of the dialogue can be distinguished in either *single* or *repeated* iterations of the dialogue as a game.

1. **Single Dialogue Game** The protocol as a single dialogue game is when the proponent's signature comprises of a single entity, therefore the mappings always found independently, as there will be no other accepted mappings in the CS. A success or failed attempt of this single dialogue game is illustrated in Figure 5.17. This independence can also be seen in a second variant of the dialogue as a single

iterated game with restrictions placed on the mappings accepted in the alignment. This variable defines the restriction on 1:1 mappings in the alignment, meaning that once a concept in O' has been mapped and accepted it can not be put forward as a candidate correspondence in a propose move.

Another variant of the single dialogue game is that a candidate mapping can be accepted without the support of a premise found in a *justify-testify* loop. Making the assumption that an argument can be posed with a claim excluding a premise, as the agents reach state S3 for the first time, an assert move will be available for the proponent agent. The traversal of the dialogue under these conditions can be seen in the Figure 5.17 (a). If this restriction is placed on the agent's strategy, a candidate mapping asserted in the dialogue can be accepted or rejected without this structural support. For the purposes of this thesis, this variant is not available to the agents, as it is defined in the protocol that the agents must have a premise to support a claim in a posed argument for a candidate mapping before it can be accepted.



(a) dialogue ending in successful concept mapping (b) dialogue ending without a concept mapping

FIGURE 5.17: Traversal of a single game dialogue, where no neighbourhood support is required in the Agent's strategy in order to accept a candidate mapping.

2. **Repeated Dialogue Game**

   In contrast to the single dialogue game the repeated dialogue game, can either iterate through a signature finding unsupported mappings as described above, or iterating through the *justify-testify* loop generating structure support of the premise of a unsupported candidate mapping. This variable seen in Figure 5.18 is defined in the agent's strategy and for the purpose of the dialogue approach, it is assumed that the agents can not accept a mapping without this support. How this support is shared however, is defined in the strategy and can be either shared a single element at a time, or sharing the full neighbourhood, where the path length is 1, related to a concept being negotiated.

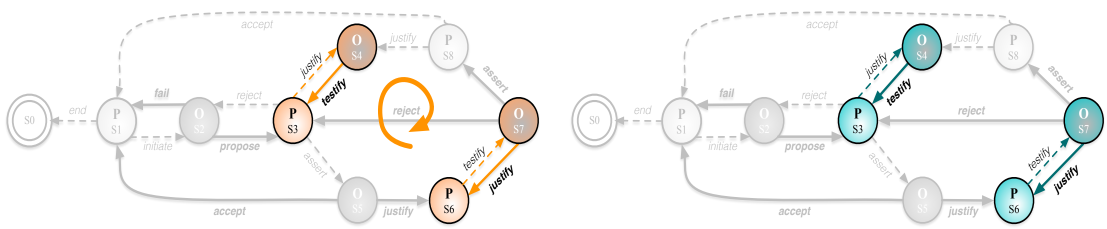
   A key definition of what determines a repeated dialogue game over a single dialogue game is centred on the proponent's signature of which is to be mapped. If this signature consists of a single entity the dialogue will only run once. However, if the signature consists of two or more concepts to be mapped the dialogue will be iterated until this is an empty set. A defining variable on how these mappings are accepted, either independently or with respects to those already stored in the CS, is defined in the mapping restrictions specified in the protocol.

Another variant of the repeated dialogue game is found within the mapping restrictions set out in the dialogue protocol. This variable defines the restriction on 1:1 mappings in the alignment meaning that once a concept in O' has been mapped and accepted, it can not be put forward as a candidate correspondence in a future *propose* move. This variant of the repeated dialogue game requires the only previously unmapped entities to be put forward in a *propose* move, which utilised the trace of the previously accepted correspondences in the CS.

In contrast to this, relaxing the protocol in order to permit 1:* mappings allows for the same entities to be proposed within different candidate mappings. In this case both dialogues will terminate as the stopping condition relies not on the mapping of the concepts in O', but on the iteration of the dialogue until the signature of the proponent is an empty set.

The versions of the approach implemented in the experiments evaluating this dialogue protocol are designed on variants of the repeated dialogue game, these variants are detailed in Chapter 8 of this thesis, however for clarity are introduced here.

The first variant of the repeated dialogue game DbMN_5, dialogue based meaning negotiation, relaxed the mapping restrictions, where 1:* mappings are accepted and the sharing in both the propose move (a single concept) and the testify move (a single triple) are done, using only one element shared in the move.



(a) Dialogue sharing support one by one. (b) Dialogue sharing support in a single multiple unit.

FIGURE 5.18: Traversal of a repeated game dialogue, where the neighbourhood support is shared in a single or multiple move.

DbMN_6 also relaxed the restriction, allowing for 1:* mappings and presents the agents the ability to share a full neighbourhood of a concept in a single *Testify* move and the neighbourhood similarity value is computed over all possibilities. This restriction means that the *Justify-Testify* loop within the dialogue is only ever performed once by the agents.

Similar to DbMN_5, DbMN_7 allows for single triple sharing however limits the mapping restrictions to 1:1 mappings, meaning that once a concept is mapped it is removed and can not be put forward as a proposed entity. This restriction is put in place to evaluate the ability of the approach to reduce the sharing of the opponent agent's ontology.

5.5 DbMN Protocol Properties

5.5.1 Pathways

the DbMN dialogue has two possible outcomes given a concept to be matched: accept or fail. Both of these dialogue outcomes lead to the dialogue termination at state S0. Either outcomes represent acceptable solutions to the alignment problem, with *fail* explicitly capturing the fact that the agents cannot find a suitable correspondence within the constraints dictated by their strategies. The conditions underlying these outcomes are described below by referring to the states in the diagram in Figure 5.4. The pathway to a *failure* in the dialogue is discussed first, and is illustrated in Figure 5.19:

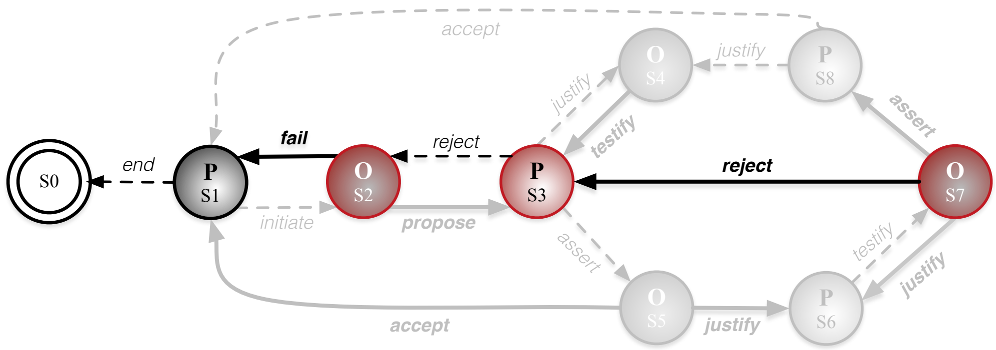


FIGURE 5.19: States and paths in the dialogue resulting in a failed outcome.

State S2: The proponent initiates the dialogue requesting a match for an entity e (state S1), however no entity e' in the opponent signature is a viable match for e, i.e. $\forall e' \in \mathsf{N_C}^{\hat{x}} \sigma_l(e, e') < \epsilon_l$.

State S3: if a candidate match for e is found in S1, the opponent responds with an entity e'. The proponent then evaluates the potential correspondence (e, e'): if this is not viable (i.e. $\sigma_l(e, e') < \epsilon_l$) then it rejects it and the dialogue fails. If the correspondence is viable then the proponent might still request the opponent to provide further evidence supporting this proposal and hence enter a *justify-testify* loop (states S3-S4). If the evidence provided is not deemed sufficient the proponent can reject the correspondence.

State S7: Following state S3, the proponent assesses the correspondence proposed: If this is found suitable it is asserted at state (S5). This assertion, however, requires some verification from the opponent who requests that the proponent provides some supporting evidence for the assertion through a further *justify-testify* loop (states S6-S7, but this time with the agents switching roles, therefore the proponent plays the role of the opponent, and vice versa). If the opponent deems that the evidence is not sufficient it will reject the assertion made by the proposer and the dialogue will fail.

The pathways for the *successful* termination of the dialogue are those in which the agents are able to agree upon a meaningful correspondence. These are identifiable are in Figure 5.20:

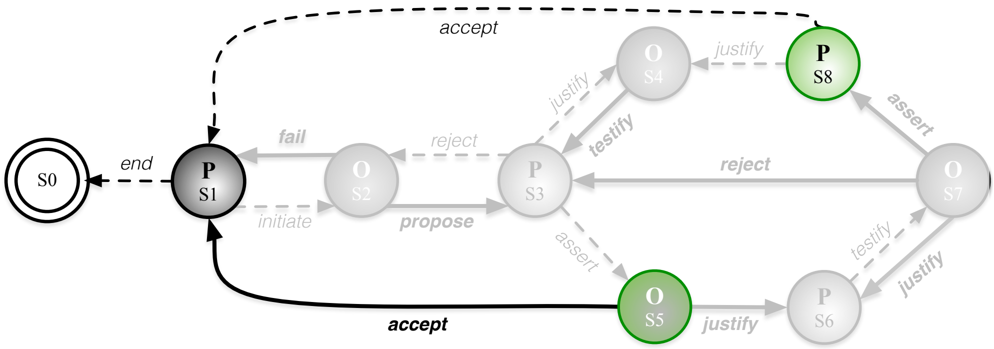


FIGURE 5.20: States and paths in the dialogue resulting in a successful outcome.

State S3: Following state S1, the opponent responds with an entity e'. The proponent then evaluates the potential correspondence (e, e') and finds that it satisfies its strategy and hence asserts the viability of the correspondence. However, the proponent may also require additional evidence from the opponent (*justify-testify* loop, states S3-S4). If the evidence is deemed sufficient then the proponent asserts the acceptability of the correspondence from their side. The correspondence is then evaluated by the opponent, who can confirm its acceptability of the correspondence with respect to its strategy (state S6) and the dialogue terminates successfully.

State S5: The opponent might also require further evidence (states S6, S7) and if satisfied, can assert the correspondence as viable from their perspective, and then this is assessed by the proposer (S8) with or without the need for further supporting evidence. If the evidence is requested, then it will be assessed and if it satisfies the agent's internal strategy then it will be accepted.

State S8: The proponent, at state S8 in the dialogue, can accept a correspondence asserted by the opponent. If the dialogue has reached this state, both agents have acceptable level of support for an argument and here the correspondence can be accepted resulting in a successful termination of the dialogue.

5.5.2 Assumptions and agent attitudes

The agents participating in this dialogue hold beliefs regarding the individual ontologies; these beliefs conflict if there are differences in the way a domain is represented. These conflicts do not need to be logical inconsistencies, but could simply be differences in the conceptualisation of the domain due to heterogeneity (as discussed in Chapter 3). For example: agents might have this statement in their ontology, $Human \sqsubseteq Livingbeing$ whereas, agent two might believe the statement $Human \sqsubseteq Mammal$. This example illustrates conflicting beliefs in the terms used and the modelling of the concept *Human*, which would need to be resolved in order to generate an accurate matching over these inconsistent terms.

The *goal* of the dialogue, as explained in Section 5.1.1 is to find a correspondence between concept c in the disclosable signature of a proponent agent $c \in \Sigma_d^x$ and c' in the disclosable signature of an opponent agent $c' \in \Sigma_d^{\hat{x}}$. This decision making process involves exchanging information about possibly conflicting knowledge. Each of the agents has internally consistent knowledge modelling an ontology O. The correspondence generation process is potentially conflicted if there are axioms in the ontology O that conflict with axioms in ontology O'.

As outlined in Chapter 4, it is assumed that the participating agents are co-operative and that they are truthful with respect to the facts they introduce throughout the DbMN dialogue, corresponding to the maxims defined by Grice and the rules of conversation defined by Pask. It is also assumed that the agents both have the common goal of reaching some form of agreement on a correspondence and have a *confident assertion attitude*, i.e. both agents will present valid arguments which they can construct with respect to satisfying their internal strategy. Agents also have a *sceptical acceptance attitude*, i.e. they will only accept correspondences for which some evidence is provided, which is developed through the notion of an argument A, having both a claim Cl, and corroborating support Pr.

Agents attitudes are used to predict the moves that an agent will perform given a protocol that defines the preconditions to moves available to an agent without specifying the actual move performed.

5.5.3 Properties

It is customary to analyse a dialogue in terms of its *soundness*, *completeness* and *termination* properties. Usually *soundness* and *completeness* are not considered in isolation but they are analysed with respect to the compliance illustrated by the dialogue, of the specific agents' strategies.

Suitable quality criteria for evaluating dialogue protocols include the *capacity* to reach an outcome regarding the termination of a dialogue and the *quality* of the outcome [4].

Completeness

Completeness is strategy dependant and relates to the assumptions and attitudes of the participating agents. Completeness is related to the notion of pre-determinism [4], i.e. the notion that under some circumstances, the result of the interaction can be established without having the dialogue itself.

If the agents have complete knowledge over the ontology and the strategy, then the outcome of any dialogue can be identified for any given entity proposed. Parsons et al [103] characterise dialogue completeness in terms of *protocol completeness* and *topic completeness*. A dialogue P is considered to be topic-complete ($AG(D)_T$) when neither agent (assuming two participants) can add anything, or construct further arguments to undercut or add to the dialogue that would change of status of the subject. D *Topic completeness* could be shown by the DbMN dialogue, modelling a graph using the Toulmin model of an argument discussed in Chapter 4, and then showing that at S0, upon the acceptance of an argument by the agents, neither participant can add any argument that challenges the support to a claim. However the formal proof is not pursued within this thesis.

The second notion of completeness defined by Parsons is that of *protocol completeness* ($AG(D)_P$), which is explicitly reliant upon the dialogue protocol P itself.

The property of protocol completeness prevents an agent A from making moves even in cases where the protocol adopted is not topic complete. Parsons states that if a dialogue is topic-complete it is also protocol-complete as if neither agent can add anything to the dialogue that would change of status of the subject, there are no legal moves that can be made to further the dialogue. Using these definitions of completeness the DbMN dialogue can be shown formally as both topic and protocol complete. Whilst this thesis does not provide a formal proof of completeness, the experimental section in Chapter 8, addresses the issues of soundness and completeness through the evaluation of precision and recall with respect to benchmark.

Soundness

A protocol is defined as *sound* following the definition in [17] once an argument has been accepted by both agents there is no argument available to the opponent that will 'attack' the proponent's proposal [17].

A *sound* dialogue protocol can be roughly restated as obtaining a 'successful' dialogue result i.e. verifying that the claim of the dialogue is 'acceptable' directly relating to the adopted strategies [42]. The dialogue detailed in Section 5.1.4 allows agents to only put forward new arguments either by proposing a new correspondence or by providing evidence supporting some candidate correspondence. Thus at the end state of the DbMN dialogue, if a correspondence is accepted, the agents have no opportunity or reason to make an 'attack'. Once arguments are uttered within the DbMN they cannot be retracted, therefore once claims and supports are uttered they will not be changed. Soundness in terms of this dialogue is addressed using an evaluation in terms of precision and recall values of the generated alignment, in comparison with a reference alignment, which is seen as a 'correct' and comparable reference alignment.

Termination

Given that the dialogue admits only two possible outcomes (either a success of failure) and it cannot propose correspondences or supporting evidence already proposed, it is possible to show that the dialogue terminates.

Proposition 5.1. *The negotiation dialogue with the set of moves $\mathcal{M}$ detailed in Section 5.2.1 will always terminate.*

Proof. Both the agents participating in the dialogue have finite disclosable ontology signatures and can only propose one entity to align at a time. Once the entity is proposed, the agents can request that the correspondence is justified in terms of its support through the entity relations within the justify and testify moves; however, this support is also finite, being bounded by the size of the disclosable signature of the ontology. Therefore in the worst case, the agents will share all the relations associated with an entity within their ontology, as the ontology is finite, the dialogue in this case will still terminate.

At any point in the dialogue, agents can only add new evidence or assert new correspondences (after having rejected a previous proposal), but are prohibited from revisiting either a correspondence or some evidence previously discussed (i.e. agents can only add to the commitment store and not retract from it). This restriction on revising correspondences also illustrates termination as once a correspondence is accepted or rejected the negotiation on that correspondence is complete and as the ontologies are finite there is a finite number of entities within the task thus, once the agents have iterated through the entities in the task the dialogue will terminate.

If the dialogue does not end before every possible viable correspondence is considered from the proponent's signature (states S1-S3), then it will end, in its worst case, once the (finite) set of *testify* - *justify* moves providing evidence for the correspondence in the claim have all been made. If no appropriate evidence is provided, then the dialogue will terminate following a *fail* outcome. □

5.6 Summary

This chapter has presented the dialogue in terms of the protocol and the agents strategies and the variants of the dialogue as a game. Here the components of the dialogue are detailed to include the assumptions that have been made within this approach, which are presented in a summary table in Table 5.5.

The dialogue protocol has been introduced in the state diagram which illustrates the model, with the available moves and the nine states of the protocol. This protocol has been detailed and formalised in terms of the moves available to the agents at the various states within the dialogue. These moves have been defined regarding their pre and postconditions, the agent who utters the move, and what information is passed in the message if any. The dialogue parameters used by the agents in order to accept or

Subject	*Assumption*
Commitment Store	A public store committing agents to statements sent from agent to agent within the dialogue.
Gamma Store	A private store storing the knowledge garnered presented as a partially connected graph.
Argument	Arguments are made by the agent to assert the validity of a correspondence.
Task	Is the set of proponent's concepts all of which need to be accepted or rejected as mapped for the dialogue to be complete and terminate.
Agents	This work assumes a maximum of two co-operative agents participating in the dialogue over a given task.
Protocol	This work assumes that the two participating agents each have an assigned ontology and that proponent and opponent roles are also assigned to the agents.
Ontologies	It is assumed that the ontologies are machine readable.
Mappings	Only equivalence mappings ($\equiv$) are considered in generating an alignment.
Labels	It is assumed that each entity shared and explored by the agents has exactly one label.

Table 5.5: Summary of assumptions made within this work

reject a candidate mapping have been defined along with the formalisms of the protocol in terms of what constitutes an argument, and the premise supporting it.

This chapter has also detailed the agents strategy it terms of the decision mechanisms that will defined which move an agent picks are a given state, and illustrated the various paths through the dialogue resulting in either a successful generation of an alignment or a failure. The key properties of the dialogue protocol which have been introduced in Chapter 4 are formally defined and detailed in this chapter, including the commitment store and the respective gamma stores used by the agents to develop a trace of moves and messages passed throughout the dialogue. This chapter has also defined the protocol properties in terms of completeness and soundness and detailing the termination conditions of the dialogue.

In Summary of this Chapter:

- Defined and detailed the pre and post conditions of the moves illustrated in the dialogue state diagram of the protocol.
- The properties required by the agents within this protocol have been formally defined including the arguments, correspondences and the requirement of a premise.
- Detailed the agent's decision mechanisms defining which move will be selected at a given state when more than one legal move is available to a sending agent.
- Illustrated the paths through the dialogue resulting in either a failed or successful outcome in generating an mutually accepted alignment.

- Outlined the proof regarding the termination of the dialogue.
- Defined the variants of the dialogue in terms of a single or repeated dialogue game, which are used in the implementation of this approach in Chapter 8.
- Detailed the dialogue properties in terms of the completeness, soundness and termination.

Chapter 6

Dialogue Walkthrough

Chapter Outline

'Alice ventured to ask...'. - Lewis Caroll

The communication and interaction between two agents in generating an alignment over their ontologies, is the central focus of this work. It is in this chapter that the protocol and strategy defined and formalised in the previous chapter is put into use by an example using the participants ***Alice*** *and* ***Bob****. This interaction is detailed move by move, illustrating the dialogical approach to ontology matching. It is essential to establish semantic meaning in order to generate an accurate alignment as no prior knowledge is shared. This approach iteratively investigates paths of an ontological knowledge base and is illustrated using academic ontology fragments for the purpose of this example.*

6.1 Protocol Review

This chapter begins with a recap of definitions for the dialogue protocol and the decision making components used by the agents, prior to detailing a walkthrough example of the dialogue used between two agents: a proponent (*Alice*) and opponent (*Bob*). This walkthrough is divided into two parts, using the proponent signature which is made up of two concepts to be mapped throughout the use of the dialogue protocol.

Moves

Firstly a review of the syntax of each move available to the agents, is summarised in Table 6.1 and is represented in the form $m = \langle x, \tau, e, e', l \rangle$, where τ is the move type such that $\tau \in \mathcal{T}$, and $\mathcal{T} = \{\mathsf{initiate}, \mathsf{propose}, \mathsf{assert}, \mathsf{accept}, \mathsf{reject}, \mathsf{testify}, \mathsf{justify}, \mathsf{fail}, \mathsf{end}\}$; e is the source entity being from the proponent; e' is the current candidate target entity; and l represents a list of zero or more additional elements (depending on the type of move). It may not be necessary to specify the source or target entity or any additional elements in which case they are represented with *nil*.

Syntax	**Description**
$\langle x, \mathsf{initiate}, e, \text{nil}, \text{nil} \rangle$	A new source entity e is proposed with the aim of finding a possible correspondence.
$\langle x, \mathsf{propose}, e, e', \text{nil} \rangle$	A new (*i.e.* not previously disclosed) candidate entity e' is proposed which lexically matches e.
$\langle x, \mathsf{justify}, e, e', \text{nil} \rangle$	A new ϖ is requested to support the candidate correspondence between e and e'.
$\langle x, \mathsf{testify}, e, e', \varpi \rangle$	If an undisclosed ϖ is known that supports the candidate correspondence (with the highest ranking predicate), then it is shared; otherwise $\varpi = \text{nil}$.
$\langle x, \mathsf{assert}, e, e', A \rangle$	The candidacy of a correspondence between e and e' is asserted, with the supporting argument A containing a subset of disclosed ϖ pairs whose aggregate *neighbourhood similarity* σ_n supports the candidacy. Note that A and σ_n are presented in Chapter 5.
$\langle x, \mathsf{accept}, e, e', A \rangle$	The candidacy is accepted if the neighbourhood similarity σ_n of the premise in A is above threshold given the sending agent's similarity metrics.
$\langle x, \mathsf{reject}, e, e', \text{nil} \rangle$	The candidacy is rejected if the neighbourhood similarity σ_n of the premise in A is below threshold given the sending agent's own similarity metrics, and no other supporting evidence is available.
$\langle x, \mathsf{fail}, e, \text{nil}, \text{nil} \rangle$	No further undisclosed candidate entities could be found that lexically match e.
$\langle x, \mathsf{end}, \text{nil}, \text{nil}, \text{nil} \rangle$	The proponent terminates the dialogue.

TABLE 6.1: A summary table outlining the set $\mathcal{T}$ of legal moves permitted by the dialogue.

Dialogue Components

The dialogue components formalised in Chapter 5 are summarised here for the purposes of this example.

Commitment Stores. The agents participating in this dialogue, manage two commitment stores. The possibility of retraction of a commitment in this work is not permitted, therefore commitment in the dialogue is permanent whilst the dialogue is open.

1. A private knowledge base managed by each agent, differentiates between the proponent's store Γ, and the opponent's store, Γ'. These gamma stores Γ are a form of private 'commitment store', logging the ontological structure of the *opponent* agent's knowledge base that has been garnered through the messages passed throughout the dialogue. The content of the private Γ stores are updated as a result of a message being sent to a receiver within the moves including:
   - *Testify* in the form of: $\langle x, \textsf{testify}, e, e', \varpi \rangle$ or $\langle \hat{x}, \textsf{testify}, e, e', \varpi \rangle$
   - *Assert* in the form of: $\langle x, \textsf{assert}, e, e', A \rangle$ or $\langle \hat{x}, \textsf{assert}, e, e', A \rangle$

   Once a message is passed to a receiver, that message is stored within the receiver's Γ, until the dialogue is terminated. It is only when the *end* is uttered at state S1 that the contents of the Γ for both agents are cleared.
2. A public knowledge base or *commitment store* is managed by both agents. Although the agents maintain individual copies of the CS, these will always be identical and contains a trace of all of the moves uttered by each agent. The contents of the private CS stores are updated as a result of a message being sent to a receiver in the move:
   - *Accept* in the form of: $\langle x, \textsf{accept}, e, e', A \rangle$ or $\langle \hat{x}, \textsf{accept}, e, e', A \rangle$

   As with the private Γ stores the CS is only cleared when the *end* is uttered at state S1 indicating a termination of the dialogue.

Arguments. Arguments are used within this dialogue in order for the agents to propose a candidate mapping. This argument is expressed as a relationship between the *claim* and the *support*, such that if the support holds then the claim must also hold. The dialogue mechanism uses arguments allowing the agents to propose candidate correspondences, and to justify them or refute them on the basis of a given fact. These facts are provided with this argument as a means of support. Agents can only make arguments that assert the validity of a new correspondence that was not previously disclosed or question its correctness by stating an alternative correspondence for one of the same entities. An argument is a pair $A = (Pr, Cl)$, where $Pr \subseteq \mathcal{L} \cup \{\top\}$ and $Cl \in \mathcal{L}$. *Args($\mathcal{L}$)* are defined as the set of all arguments derivable from the language $\mathcal{L}$. In this definition Pr is the *support* (representing a set of premises of an argument), whilst Cl is the *claim*. Facts (*i.e.* statements with no premises) are represented as $(\top, Cl)$. As each new argument either introduces a new correspondence, or states a new premise for an existing one, there is no possibility of cycles in arguments and thus the agents will either reach an agreement or they will reject the proposal.

Lexical similarity metric. This is the function $\sigma_l : \mathsf{N_C} \times \mathsf{N_C} \rightarrow [0,1]$ which returns the lexical similarity between the labels of two entity names $e, e' \in \mathsf{N_C}$, such that $\sigma_l(e, e') = 1$ iff $e = e'$ and 0 if the two labels are different. This function is used in the *initiate* move of the dialogue to discover those entities in agent O's signature (Alice) that could lexically match an entity in agent O' ontology (Bob). A lexical match is considered *viable* if $\sigma_l(e, e')$ is greater or equal to its threshold ϵ_l.

Structural similarity metric. This is the function $\sigma_s : \Pi \times \Pi \rightarrow [0,1]$ that returns the structural similarity between two triples $\varpi, \varpi' \in \Pi$, such that $\sigma_s(\varpi, \varpi') = 1$ if the two triples are considered as equivalent and 0 otherwise. As *subject-predicate-object* triples relating to e' are disclosed by one agent, the second agent should try to identify similar localised structures in its own ontology. This may be based purely on the triples themselves or may also take into account other information that has so far been ascertained or inferred. As with the σ_l function, this example make no assumptions about how the similarity function is defined and has generated these values arbitrarily for the purposes of illustrating the protocol.

Neighbourhood similarity. This is the function $\sigma_n : \{(\varpi, \varpi') \in \Pi \times \Pi \mid \varpi \in \Gamma, \varpi' \in O\} \rightarrow [0,1]$ that returns an aggregate similarity calculated from a matching generated by calculating all possible structural similarities between the triples in an agent's gamma store Γ and the triples in the disclosable fragment of the opponent's ontology O', such that $\sigma_n(\varpi, \varpi') = 1$ if the neighbourhood is structurally equivalent and 0 otherwise. This σ_n is computed over the set of all matching (ϖ, ϖ') *pairs*, such that no triple from one ontology is 'paired' to more than one triple in the other ontology, restricting a *one-to-one* mapping between the sets of triples. The premise Pr for the claim proposed by agent for a correspondence c will comprise a subset of pairs from the set $\mathsf{pairing}(\Gamma, O)$, with a corresponding aggregate *neighbourhood similarity* σ_n. For the purposes of this example the σ_n is based on the structural similarity scores σ_s for each triple pair and is illustrated in Table 6.2.

Rank function. This property of the dialogue determines the order of sharing the agents adopt regarding the structural details about the ontology in the neighbourhood of an entity under consideration. Agents request triples belonging to the local neighbourhood of some entity e' in the other agent's ontology to support the candidacy of a correspondence. Therefore it is assumed that each agent utilises a function $\mathsf{rank}(e)$ that generates a strict pre-ordering of triples for a given subject e. This is formally defined as: The **rank function**, $\mathsf{rank} : \mathsf{N_C} \rightarrow \mathcal{R} \subseteq \Pi$ returns an ordered list of triples in a path starting at some entity $e \in \mathsf{N_C}$, where $\forall \varpi_i, \varpi_j \in \mathcal{R} : \varpi_i \succcurlyeq \varpi_j$. For the purposes of this example a ranking for the agents has been developed and is presented in Section 6.2

6.2 Dialogue Walkthrough

By means of an example the dialogue protocol illustrates how two agents find an alignment between the public signatures of their ontologies. Two agents *Alice* and *Bob*, each possess a private ontological fragment seen in Figure 6.1.

The dialogue between the agents, *Alice* and *Bob* is undertaken over a signature of concepts, the proponent agent has, which requires mapping. This signature, for the purpose of this example, is developed by the proponent *Alice*, and consists of the concepts:

$$\Sigma^t = \{\mathit{Author}, \mathit{Paper}\},$$

This signature defines the concepts used in the initiate move of the dialogue. Here *Alice* requires both the concepts *Author* and *Paper* for the dialogue to be seen as complete.

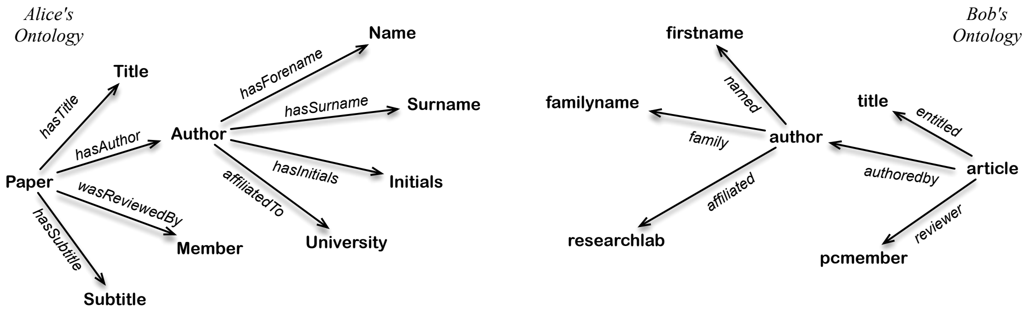


FIGURE 6.1: Two trivial ontology fragments for *Alice* and *Bob* used in the walkthrough example.

Both agents implement different structural similarity metrics σ_s, and a subset[1] of the values for different ϖ triple pairs are used within this example and given in Table 6.2. For example the structural similarity[2] σ_s between the triple $\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title} \rangle$ and $\langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle$ for *Alice*, $\sigma_s^{Alice} = 0.70$, whereas for *Bob* the similarity for this pair is $\sigma_s^{Bob} = 0.68$.

Alongside the structural similarity metrics, both agents generate a strict pre-ordering of the properties for each entity e, using the function $\mathsf{rank}()$. For the purposes of this first part of the example in relation to the initiating concept *Author* the pre-orderings are:

$$\mathsf{rank}^{Alice}(\mathit{Author}) = \{\mathit{hasSurname}, \mathit{hasForename}, \mathit{affiliatedTo}, \mathit{hasInitials}\}$$
$$\mathsf{rank}^{Bob}(\mathit{author}) = \{\mathit{family}, \mathit{named}, \mathit{affiliated}\}$$

It is assumed that a neighbourhood similarity metric $\sigma_n(Pr)$ calculates the average structural similarity $\bar{\sigma}_s$ of the triple pairs in the premise Pr, with a coefficient that

[1]Although other similarity pairs have been calculated, these do not appear in the dialogue example (for example, because the distance is lower than those explicitly stated), and thus have not been given for brevity.

[2]These similarity pairs are not generated a priori but are calculated during the dialogue.

increases asymptotically as the cardinality of Pr increases. The metric is defined as $\sigma_n(Pr) = \bar{\sigma}_s \times (1 - \frac{1}{2(|Pr|+1)})$. Both a neighbourhood threshold $\epsilon_n = 0.55$ and a lexical threshold $\epsilon_l = 0.75$ are also assumed for this example. These values have been selected through manual calculation, to best represent the behaviour of this dialogue approach for the purpose of this example.

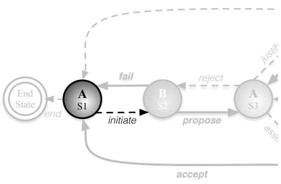
6.2.1 First iteration of example

The example dialogue is presented in two parts and summarised in Tables 6.3, 6.4 and 6.5 and is divided in terms of the two concepts in the proponent's signature Σ^t. Firstly an example of the beginning of a dialogue at state S1 in Figure 5.4 of the state transition diagram. The second part of the example begins directly after the first in assuming that an alignment between concepts has already been generated. The dialogue begins at state S1 with a second *initiate* move from the proponent agent. The end of this second example marks the end of the dialogue and terminates in state S0.

Alice's ϖ	*Bob*'s ϖ	σ_s^{Alice}	σ_s^{Bob}
⟨*Author*, *hasSurname*, *Surname*⟩	⟨*author*, *family*, *familyname*⟩	0.63	0.60
⟨*Author*, *affiliatedTo*, *University*⟩	⟨*author*, *affiliated*, *researchlab*⟩	0.85	0.86
⟨*Author*, *hasInitials*, *Initials*⟩	⟨*author*, *named*, *firstname*⟩	0.67	0.83
⟨*Author*, *hasForename*, *Name*⟩	⟨*author*, *named*, *firstname*⟩	0.71	0.69
⟨*Paper*, *hasTitle*, *Title*⟩	⟨*article*, *entitled*, *title*⟩	0.70	0.68
⟨*Paper*, *hasAuthor*, *Author*⟩	⟨*article*, *authoredby*, *author*⟩	0.65	0.61
⟨*Paper*, *hasSubtitle*, *Subtitle*⟩	⟨*article*, *entitled*, *title*⟩	0.68	0.84
⟨*Paper*, *wasReviewedBy*, *Member*⟩	⟨*article*, *reviewer*, *pcmember*⟩	0.66	0.60

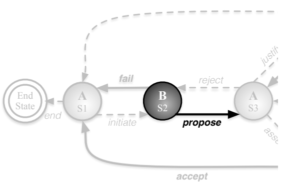
TABLE 6.2: The structural similarities of possible corresponding triples between *Alice* & *Bob*'s ontologies. Whilst not exhaustive lists a subset of triples between the two ontologies.

Move 1: State *S1*, *initiate* move.



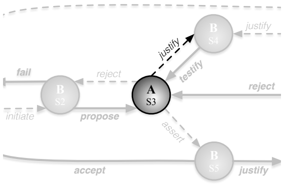
Alice utters an *initiate* move. Here Alice states that she wants to discuss a possible match for the entity *Author* in her ontology. The nature of the selection of this initial concept is extraneous and therefore can be selected at random from the concepts within the Σ^t to be mapped.

Move 2: State *S2*, *propose* move.



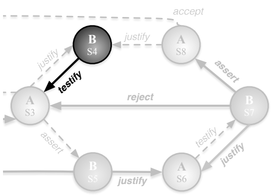
Bob identifies *author* as the most similar entity in his ontology to the entity *Author* with a lexical similarity $\sigma_l^{Bob}(\mathsf{Author}, \mathsf{author}) = 0.86$ (this value is not given in the table). As this is above threshold ϵ_l, he can then respond with the move ⟨*Bob*, *propose*, *Author*, *author*, nil⟩.

Move 3: State *S3*, *justify* move.



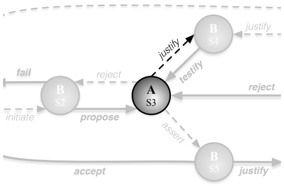
Alice now knows that $\langle \mathit{Author}, \mathit{author}, \equiv \rangle$ is a potential correspondence c (based on *Bob's* lexical similarity claim). She verifies that her lexical similarity for the entity pair is above threshold (in this case $\sigma_l^{Alice}(\mathit{Author}, \mathit{author}) = 0.76$), and then asks *Bob* to provide some evidence to justify the candidacy of c. At this point neither agent has support for c; i.e. $Pr = \varnothing$.

Move 4: State *S4*, *testify* move.



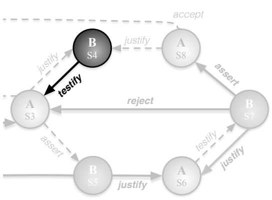
Bob (state S4) determines the next property that has *author* as its *domain* and that has not yet been disclosed and therefore has not yet appeared in the commitment store CS. Given that his ranking for the entity *author* is $\mathsf{rank}^{Bob}(\mathit{author}) = \{\mathit{family}, \mathit{named}, \mathit{affiliated}\}$ and that so far none of these properties have been disclosed, he shares the fact that the highest ranked property *family* relates the two entities *author* and *familyname* in his ontology with the triple $\langle \mathit{author}, \mathit{family}, \mathit{familyname} \rangle$.

Move 5: State *S3*, *justify* move.



Alice tries to determine if there is sufficient support for c. She realises that $\langle \mathit{Author}, \mathit{hasSurname}, \mathit{Surname} \rangle$ in her ontology is the most similar triple to the one *Bob* disclosed in move 4, with a similarity $\sigma_s^{Alice} = 0.63$ (Table 6.2). She calculates that the premise $Pr = \{(\langle \mathit{Author}, \mathit{hasSurname}, \mathit{Surname} \rangle, \langle \mathit{author}, \mathit{family}, \mathit{familyname} \rangle)\}$ has a neighbourhood similarity $\sigma_n^{Alice} = 0.63 \times (1 - \frac{1}{2(|Pr|+1)}) = 0.63 * 0.75 = 0.473$. She will only *assert* an argument for c if this is above the threshold $\epsilon_n = 0.55$. As this is below threshold, she requests additional evidence to justify c.

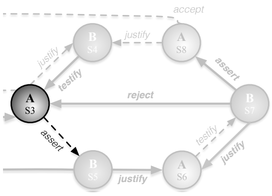
Move 6: State *S4*, *testify* move.



Bob's next highest ranked property that has not been disclosed (*i.e.* does not appear in CS) whose domain is *author*, is the entity *named*. Therefore he shares the triple $\langle \mathit{author}, \mathit{named}, \mathit{firstname} \rangle$.

Move 7: State *S3*, *assert* move.

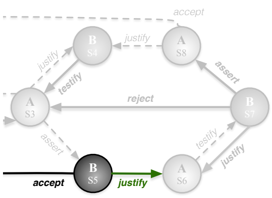
Alice firstly verifies if one of her triples is similar to that disclosed by *Bob* in move 6.



Although she has two triples that share their highest similarity with *Bob's* disclosed triple, she chooses $\langle \mathsf{Author}, \mathsf{hasForename}, \mathsf{Name} \rangle$ as the similarity is higher than $\langle \mathsf{Author}, \mathsf{hasInitials}, \mathsf{Initials} \rangle$. She adds this to Pr and calculates the neighbourhood similarity $\sigma_n^{Alice} = (0.63 + 0.71)/2 \times (1 - \frac{1}{2(2+1)}) = 0.67 * 0.8\dot{3} = 0.556$, which (from *Alice's* perspective) is above threshold, Therefore she proposes the argument A for the correspondence $c = \langle \mathsf{Author}, \mathsf{author}, \equiv \rangle$, given that:

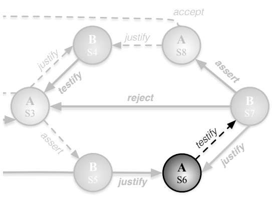
$$Pr = \{(\langle \mathsf{Author}, \mathsf{hasSurname}, \mathsf{Surname} \rangle, \langle \mathsf{author}, \mathsf{family}, \mathsf{familyname} \rangle), \\ (\langle \mathsf{Author}, \mathsf{hasForename}, \mathsf{Name} \rangle, \langle \mathsf{author}, \mathsf{named}, \mathsf{firstname} \rangle)\}$$

Move 8: State *S5*, *justify* move.



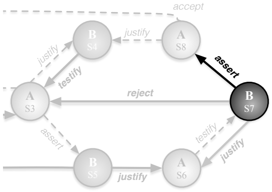
Given the argument A for the correspondence $c = \langle \mathsf{Author}, \mathsf{author}, \equiv \rangle$ which has been asserted by *Alice* in the previous move, *Bob* at (state S5) can make one of two possible moves: (i) *Bob* can *accept* the argument A if $\sigma_n^{Bob}(Pr)$ is above threshold and transition to state S1. Or (ii) *Bob* can *justify* the candidacy of c by requesting further support (if other undisclosed properties exist). In this case, *Bob* calculates that the neighbourhood similarity (from his perspective) is $\sigma_n^{Bob} = (0.60 + 0.69)/2 \times (1 - \frac{1}{2(2+1)}) = 0.65 * 0.8\dot{3} = 0.54$, which is below threshold. Therefore *Bob* asks *Alice* if she could provide some further evidence to justify c.

Move 9: State *S6*, *testify* move.



Alice then shares the triple $\langle \mathsf{Author}, \mathsf{affiliatedTo}, \mathsf{University} \rangle$ as *affiliatedTo* is her highest ranked non-disclosed property for the domain entity *Author*.

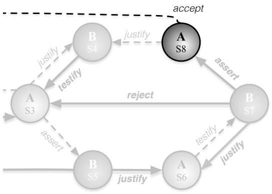
Move 10: State *S7*, *assert* move.



Bob recalculates the mean similarity for the new support to include the mapping of the triple shared by *Alice* in move 9 to his triple $\langle \mathsf{author}, \mathsf{affiliated}, \mathsf{researchlab} \rangle$: $\sigma_n^{Bob} = (0.60 + 0.69 + 0.86)/3 \times (1 - \frac{1}{2(3+1)}) = 0.72 * 0.875 = 0.627$, which is above threshold. *Bob* is therefore happy to accept the candidacy of c. It is now his turn to *assert* the new argument for c given the new premise Pr.

Move 11: State *S8*, *accept* move.

Alice confirms that from her perspective, σ_n^{Alice} is above threshold;



$\sigma_n^{Alice} = (0.63+0.71+0.85)/3 \times (1 - \frac{1}{2(3+1)}) = 0.73 * 0.875 =$ 0.639, which is above threshold, and accepts the argument with the move *accept*. This *accept* move marks the end of the first part of this example. At this point, through co-operation the agents were able to engage in the joint activity of determining a correspondence $\langle \mathit{Author}, \mathit{author}, \equiv \rangle$ between two entities *Author* and *author* based on the similarity of the local neighbourhood of these entities.

This example is summarised in Table 6.3 and demonstrates the following triples from *Bob's* ontology were shared in order to find an mutually accepted alignment for the first concept in Σ^t:

$$\langle \mathit{author}, \mathit{family}, \mathit{familyname} \rangle, \langle \mathit{author}, \mathit{named}, \mathit{firstname} \rangle, \langle \mathit{author}, \mathit{affiliated}, \mathit{researchlab} \rangle$$

The triples shared from *Alice's* ontology for this first concept in the Σ^t include:

$$\langle \mathit{Author}, \mathit{hasSurname}, \mathit{Surname} \rangle, \langle \mathit{Author}, \mathit{hasForename}, \mathit{Name} \rangle, \langle \mathit{Author}, \mathit{affiliatedTo}, \mathit{University} \rangle$$

Although all of *Bob's* ϖ triples were disclosed where the subject of the triple is the concept *author*, *Alice* was able to reach the consensus without revealing knowledge of one of her triples: $\langle \mathit{Author}, \mathit{hasInitials}, \mathit{Initials} \rangle$, even though from *Bob's* perspective it was actually more similar to *Bob's* triple $\langle \mathit{author}, \mathit{named}, \mathit{firstname} \rangle$ than $\langle \mathit{Author}, \mathit{hasForename}, \mathit{Name} \rangle$. If in move 7 *Alice* had found that the triple with the highest similarity to *Bob's* triple $\langle \mathit{author}, \mathit{named}, \mathit{firstname} \rangle$ was actually $\langle \mathit{Author}, \mathit{hasForename}, \mathit{Name} \rangle$, then *Bob* would have accepted the support in move 8 (as $\sigma_n^{Bob} = (0.6 + 0.83)/2 \times (1 - \frac{1}{2(2+1)}) = 0.72 * 0.8\dot{3} =$ 0.593, which was above threshold) and fewer properties would have been disclosed.

Walkthrough example Part I *Summary*

In summary of this part of the example, the dialogue exchange in steps 1-11, between *Alice* and *Bob* is complete for the initial concept in the Σ^t *Author*, as a mutually accepted mapping has been found between the agents. However, the dialogue can not terminate here as an alignment for *Paper* is still to be found between the two agents. It can be summarised at this point that the agents return to state S1 of the dialogue with the mapping $\langle \mathit{Author}, \mathit{author}, \equiv \rangle$ stored in the public joint commitment store CS. The dialogue will then begin again at state S1 to exchange messages in order to find a mapping for the last concept in the Σ^t, *Paper*, this mapping is investigated within Part II of this example.

Move	*Message*	Γ^{Alice}	Γ^{Bob}	*CS*
1	⟨*Alice*, initiate, *Author*, nil, nil⟩	-	-	-
2	⟨*Bob*, propose, *Author*, *author*, nil⟩	-	-	-
3	⟨*Alice*, justify, *Author*, *author*, nil⟩	-	-	-
4	⟨*Bob*, testify, *Author*, *author*, ⟨*author*, *family*, *familyname*⟩⟩	⟨*author*, *family*, *familyname*⟩	-	-
5	⟨*Alice*, justify, *Author*, *author*, ⟨*author*, *family*, *familyname*⟩⟩	⟨*author*, *family*, *familyname*⟩	-	-
6	⟨*Bob*, testify, *Author*, *author*, ⟨*author*, *named*, *firstname*⟩⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩	- -	-
7	⟨*Alice*, assert, *Author*, *author*, {(⟨*Author*, *hasSurname*, *Surname*⟩, ⟨*author*, *family*, *familyname*⟩, ⟨*Author*, *hasForename*, *Name*⟩, ⟨*author*, *named*, *firstname*⟩))}, ⟨*Author*, *author*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩	-
8	⟨*Bob*, justify, *Author*, *author*, nil⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩	-
9	⟨*Alice*, testify, *Author*, *author*, ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	-
10	⟨*Bob*, assert, *Author*, {(⟨*Author*, *hasSurname*, *Surname*⟩, ⟨*author*, *family*, *familyname*⟩, ⟨*Author*, *hasForename*, *Name*⟩, ⟨*author*, *named*, *firstname*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩, ⟨*author*, *affiliated*, *researchlab*⟩))}, ⟨*Author*, *author*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	-
11	⟨*Alice*, accept, *Author*, *author*, {(⟨*Author*, *hasSurname*, *Surname*⟩, ⟨*author*, *family*, *familyname*⟩, ⟨*Author*, *hasForename*, *Name*⟩, ⟨*author*, *named*, *firstname*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩, ⟨*author*, *affiliated*, *researchlab*⟩))}, ⟨*Author*, *author*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩

TABLE 6.3: Showing the update of *Alice* and *Bob* private and public knowledge stores throughout the example dialogue run, from move 11-23

6.2.2 Second iteration of example

In the example dialogue (Table 6.3) it is assumed that the dialogue has already commenced resulting in *Alice* accepting the correspondence $\langle \mathit{Author}, \mathit{author}, \equiv \rangle$ in a previous negotiation round (moves 1-11; the acceptance of this correspondence is illustrated in move 11 of Table 6.3). The order in which the dialogue proponent selects entities for exploration is strategic[3]; for this example it is assumed that the first two entities *Alice* explores are (in order): *Author* and *Paper*.

For the beginning of this iteration of the dialogue the metrics defined in the first part of this example still hold and can be recapped as follows:

Proponent Signature: The signature, removing all the previously mapped concepts includes: $\Sigma^t = \{\mathit{Paper}\}$

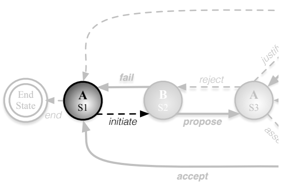
Structural similarity metrics: Both agents implement the same structural similarity metrics σ_s, as seen in the first iteration of this walkthrough example, represented in Table 6.2. For example the structural similarity σ_s between the triple $\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title} \rangle$ and $\langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle$ for *Alice*, $\sigma_s^{Alice} = 0.70$, whereas for *Bob* the similarity for this pair is $\sigma_s^{Bob} = 0.68$.

Ranking: The strict pre-ordering of the properties for each entity e, using the function $\mathsf{rank}()$, for the purposes of this second part of the example, in relation to the initiating concept *Paper* are:

$$\mathsf{rank}^{Alice}(\mathit{Paper}) = \{\mathit{hasTitle}, \mathit{hasAuthor}, \mathit{hasSubtitle}, \mathit{wasReviewedBy}\}$$
$$\mathsf{rank}^{Bob}(\mathit{article}) = \{\mathit{reviewer}, \mathit{entitled}, \mathit{authoredby}\}$$

Neighbourhood similarity metric: It is assumed that the same neighbourhood similarity metric $\sigma_n(Pr)$, as introduced in the first iteration, calculates the average structural similarity $\bar{\sigma}_s$ of the triple pairs in the premise Pr, with a coefficient that increases asymptotically as the cardinality of Pr increases. The metric is defined as $\sigma_n(Pr) = \bar{\sigma}_s \times (1 - \frac{1}{2(|Pr|+1)})$. Both the same neighbourhood threshold $\epsilon_n = 0.55$ and a lexical threshold $\epsilon_l = 0.75$ are assumed for this example to that of the first iteration.

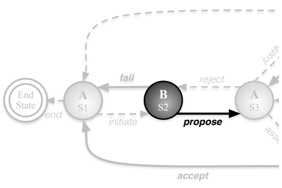
Move 12: State *S1*, *initiate* move.



It is here that the dialogue between *Alice* and *Bob* is resumed at state S1 in move 12 where *Alice* utters an *initiate* move. Having previously accepted a correspondence for *Author* (Move 11), *Alice* utters an *initiate* move (state S1 in Fig 5.4), to explore a possible correspondence for the next entity from her public signature that she wants to align; which in this case is *Paper*.

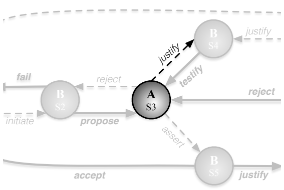
[3]As mentioned previously it is not specified here how the strategic choices are made by each agent, but assume some objective function that determines these choices exists.

Move 13: State *S2*, *propose* move.



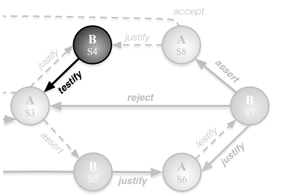
Bob identifies that the entity *article* is the most similar entity in his ontology to *Paper*, which has a lexical similarity of $\sigma_l^{Bob}(\mathit{Paper}, \mathit{article}) = 0.82$ (this value is not given in the table). As this is above threshold ϵ_l, he responds with the move $\langle Bob, \mathit{propose}, \mathit{Paper}, \mathit{article}, \text{nil}\rangle$.

Move 14: State *S3*, *justify* move.



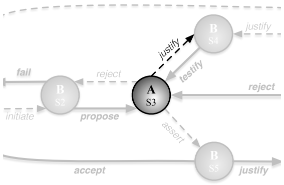
Alice now knows that $\langle \mathit{Paper}, \mathit{article}, \equiv\rangle$ is a potential correspondence c (based on *Bob's* lexical similarity claim). She verifies that her lexical similarity for the entity pair is above threshold (in this case $\sigma_l^{Alice}(\mathit{Paper}, \mathit{article}) = 0.79$). As she is aware that the entity *Paper* has a local neighbourhood (i.e. there is at least one ϖ that has *Paper* as its subject), she asks *Bob* to provide some evidence to justify the candidacy of c. At this point, neither agents have support for c; i.e. $Pr = \varnothing$.

Move 15: State *S4*, *testify* move.



Bob at (state S4) generates a strict pre-ordering of the properties for *article* using the function $\mathsf{rank}()$; i.e. $\mathsf{rank}^{Bob}(\mathit{article}) = \{\mathit{reviewer}, \mathit{entitled}, \mathit{authoredby}\}$. He uses this to determine the next property that has *article* as its *domain* and that has not yet been disclosed (i.e. that has not yet appeared in the commitment store CS). As none of the properties in $\mathsf{rank}^{Bob}(\mathit{article})$ have yet been disclosed, he shares the fact that the highest ranked property *reviewer* relates the two entities *article* and *pcmember*.

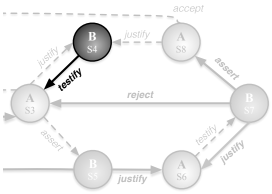
Move 16: State *S3*, *justify* move.



Alice tries to determine if there is sufficient support for c. She realises that $\langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member}\rangle$ in her ontology is the most similar triple to the one *Bob* disclosed in move 15, with a similarity $\sigma_s^{Alice} = 0.66$ (Table 6.2). She calculates that the premise $Pr = \{(\langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member}\rangle, \langle \mathit{article}, \mathit{reviewer}, \mathit{pcmember}\rangle)\}$ has a neighbourhood similarity $\sigma_n^{Alice} = 0.66 \times (1 - \frac{1}{2(|Pr|+1)}) = 0.66 * 0.75 = 0.495$. She will only *assert* an argument for c if this is above the threshold $\epsilon_n = 0.55$. As this is below threshold, she requests additional evidence to justify c.

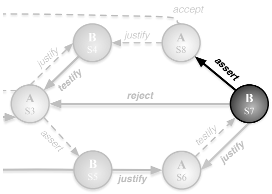
Move 17: State *S4*, *testify* move.

Bob's then finds the next highest ranked property that has not been disclosed (*i.e.* a property that does not already appear in CS).



This property also needs to have the domain *article*. In *Bob's* ontology this is the entity *entitled*. Therefore he shares the following triple ⟨*article*, *entitled*, *title*⟩ in a *testify* move.

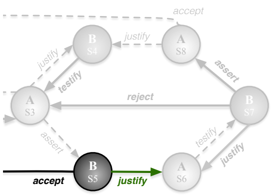
Move 18: State *S7*, *assert* move.



Alice checks to see if one of her triples is similar to that disclosed by *Bob* in move 17. Although she has two triples that share their highest similarity with *Bob's* disclosed triple, she chooses ⟨*Paper*, *hasTitle*, *Title*⟩ as the similarity is higher than ⟨*Paper*, *hasSubtitle*, *Subtitle*⟩. She adds this to Pr and calculates the neighbourhood similarity $\sigma_n^{Alice} = (0.66 + 0.7)/2 \times (1 - \frac{1}{2(2+1)}) = 0.68 * 0.8\dot{3} = 0.56$, which (from *Alice's* perspective) is above threshold, Therefore she proposes the argument A for the correspondence $c = \langle \mathit{Paper}, \mathit{article}, \equiv \rangle$, given that:

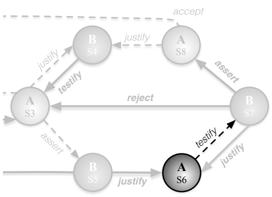
$$Pr = \{(\langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member} \rangle, \langle \mathit{article}, \mathit{reviewer}, \mathit{pcmember} \rangle),$$
$$(\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title} \rangle, \langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle)\}$$

Move 19: State *S5*, *justify* move.



Given the argument A for the correspondence c asserted in the previous move, *Bob* (state S5) can make one of two possible moves: (i) *accept* the argument A if $\sigma_n^{Bob}(Pr)$ is above threshold, and transition to state S1. Or (ii) *justify* the candidacy of c by requesting further support (if other undisclosed properties exist). In this case, *Bob* calculates that the neighbourhood similarity (from his perspective) is $\sigma_n^{Bob} = (0.60 + 0.68)/2 \times (1 - \frac{1}{2(2+1)}) = 0.64 * 0.8\dot{3} = 0.5\dot{3}$, which is below threshold. However, *Bob* is aware of other triples for the entity *article* that do not appear in Pr, and thus asks *Alice* if she could provide some further evidence to justify c.

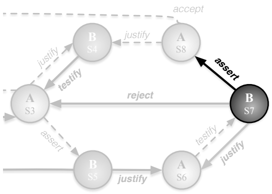
Move 20: State *S6*, *testify* move.



Alice now generates her own strict pre-ordering of the properties for *Paper*, using the function rank(); i.e. $\mathsf{rank}^{Alice}(\mathit{Paper}) = \{\mathit{hasTitle}, \mathit{hasAuthor}, \mathit{hasSubtitle}, \mathit{wasReviewedBy}\}$. She shares the triple ⟨*Paper*, *hasAuthor*, *Author*⟩ as *hasAuthor* is her highest ranked, non-disclosed property for the domain entity *Paper*. The property *hasTitle* was ranked higher however this was disclosed in her previous *assert* move.

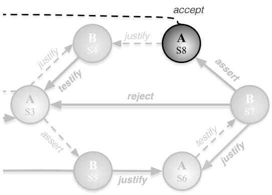
Move 21: State *S7*, *assert* move.

Bob must then recalculate the mean similarity for this support.



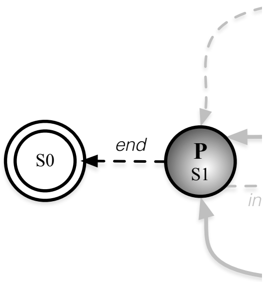
This recalculation must be inclusive of the triple shared by *Alice* in move 20: $\sigma_n^{Bob} = (0.60 + 0.68 + 0.61)/3 \times (1 - \frac{1}{2(3+1)}) = 0.63 * 0.875 = 0.551$, which is above threshold. *Bob* is happy to accept the candidacy of c. It is now his turn to *assert* the new argument for c given the new premise Pr.

Move 22: State *S8*, *accept* move.



Alice confirms that from her perspective, $\sigma_n^{Alice} = (0.66 + 0.7 + 0.65)/3 \times (1 - \frac{1}{2(3+1)}) = 0.67 * 0.875 = 0.59$, which is above threshold and she can then accept the argument.

Move 23: State *S1*, *end* move.



At this point through co-operation between *Alice* and *Bob*, the agents were able to engage in the joint activity of determining a correspondence $\langle \mathsf{Paper}, \mathsf{article}, \equiv \rangle$ between two entities *Paper* and *article* based on the similarity of the local neighbourhood of these entities. At this point all the concepts in the signature Σ^t have been mapped and this is now an empty set. Therefore the dialogue at this state S1 can now terminate.

With reference to this example illustrated in Table 6.4 the following triples from *Bob's* ontology were shared in order to find a mutually accepted alignment for the signature Σ^t:

$$\langle \mathit{author}, \mathit{family}, \mathit{familyname} \rangle, \langle \mathit{author}, \mathit{named}, \mathit{firstname} \rangle, \langle \mathit{author}, \mathit{affiliated}, \mathit{researchlab} \rangle,$$
$$\langle \mathit{article}, \mathit{reviewer}, \mathit{pcmember} \rangle, \langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle, \langle \mathit{article}, \mathit{authoredby}, \mathit{author} \rangle$$

The triples shared from *Alice's* ontology, for this first concept in the Σ^t include:

$$\langle \mathit{Author}, \mathit{hasSurname}, \mathit{Surname} \rangle, \langle \mathit{Author}, \mathit{hasForename}, \mathit{Name} \rangle,$$
$$\langle \mathit{Author}, \mathit{affiliatedTo}, \mathit{University} \rangle, \langle \mathit{Paper}, \mathit{wasReviewedBy}, \mathit{Member} \rangle, \langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title} \rangle,$$
$$\langle \mathit{Paper}, \mathit{hasAuthor}, \mathit{Author} \rangle$$

Although all of *Bob's* triples were disclosed, *Alice* was able to reach the consensus without revealing knowledge of one of her triples: $\langle \mathit{Paper}, \mathit{hasSubtitle}, \mathit{Subtitle} \rangle$, even though from *Bob's* perspective, it was actually more similar to the triple in *Bob's* ontology: $\langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle$ than $\langle \mathit{Paper}, \mathit{hasTitle}, \mathit{Title} \rangle$. If, in move 18, *Alice* had found that the triple with the highest similarity to $\langle \mathit{article}, \mathit{entitled}, \mathit{title} \rangle$ was actually $\langle \mathit{Paper}, \mathit{hasSubtitle}, \mathit{Subtitle} \rangle$, then *Bob* would have accepted the support in move 19 (as $\sigma_n^{Bob} = (0.6 + 0.84)/2 \times (1 - \frac{1}{2(2+1)}) = 0.67 * 0.8\dot{3} = 0.56$, which was above threshold) and fewer properties would have been disclosed.

Summary of Part II of the walkthrough example.

In summary of this second part of the example presented, the dialogue exchange in steps 12-23, between *Alice* and *Bob* is complete for the final concept in the Σ^t *Paper*, as a mutually accepted mapping has been found between the agents. At the point of *accept* at state S1 for the concept *Paper*, the full proponent's signature Σ^t consisting of the concepts *Author* and *Paper* is now an empty set. The dialogue can terminate here as an alignment for all the concepts in the signature has been found. It can be summarised at this point that the agents return to state S1 of the dialogue with the mapping $\langle \mathit{Author}, \mathit{author}, \equiv \rangle$ stored in the public joint commitment store CS. The dialogue will terminate at state S1 with the *end* locution uttered by the proponent agent (*Alice*), and both the public and private stores shared by the agents throughout the dialogue will be cleared.

This utterance of the *end* marks the end of the walkthrough example of the dialogue approach which is evaluated in a series of experiments. The evaluation of this approach is outlined with the experiment generalities detailing the processes and variables that occur throughout the various versions of the approach. These generalities are followed by detailed results and evaluations of the altered versions of the approach and the findings from their performance are compared with each other and current alignment systems.

This chapter has presented a detailed walkthrough example of the dialogue protocol presented up to this chapter of the thesis. This example has been broken down into two parts, illustrating the messages exchanged within the dialogue for agents *Alice* and *Bob* for the signature Σ^t { *Paper*, *Author*}, in terms of the dialogue beginning with an initiated concept, and concluding in the dialogue closing after a final match has been found. Here the participating agents, *Alice* and *Bob*, have two ontology fragments representing heterogeneous knowledge from an academic domain and utilise the dialogue protocol in order to generate a mutually agreed upon alignment.

In Summary of this Chapter:

- Summarised the protocol and the moves available to the agents.
- Presented a practical representation of the dialogue protocol and its metrics for the purposes of an illustrated example, showing the protocol in use.
- Illustrated a detailed walkthrough of the initiating concept in a signature to be mapped and concluding with this being added to an alignment agreed upon by both agents with supporting structural knowledge regarding a premise.
- Illustrated a second detailed walkthrough on a second concept in a signature concluding with the dialogue closing after the final concept is matched.

Move	*Message*	Γ^{Alice}	Γ^{Bob}	CS
11	⟨*Alice*, accept, *Author*, *author*, {(⟨*Author*, *hasSurname*, *Surname*⟩, ⟨*author*, *family*, *familyname*⟩, ⟨*Author*, *hasForename*, *Name*⟩, ⟨*author*, *named*, *firstname*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩, ⟨*author*, *affiliated*, *researchlab*⟩)⟩}, ⟨*Author*, *author*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
12	⟨*Alice*, initiate, Paper, nil, nil⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
13	⟨*Bob*, propose, Paper, *article*, nil⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
14	⟨*Alice*, *justify*, Paper, *article*, nil⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
15	⟨*Bob*, testify, Paper, *article*, ⟨*article*, *reviewer*, *pcmember*⟩⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
16	⟨*Alice*, *justify*, Paper, *article*, nil⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
17	⟨*Bob*, testify, Paper, *article*, ⟨*article*, *entitled*, *title*⟩⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩

TABLE 6.4: Showing the update of *Alice* and *Bob* private and public knowledge stores throughout the example dialogue run from move 11-23.

Move	*Message*	Γ^{Alice}	Γ^{Bob}	*CS*
18	⟨*Alice*, assert, *Paper*, *article*, {(⟨*Paper*, *wasReviewedBy*, *Member*⟩, ⟨*article*, *reviewer*, *pcmember*⟩) (⟨*Paper*, *hasTitle*, *Title*⟩, ⟨*article*, *entitled*, *title*⟩)} ⟨*Paper*, *article*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩	⟨*Author*, *author*, ≡⟩
19	⟨*Bob*, justify, *Paper*, *article*, nil	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩ ⟨*Paper*, *wasReviewedBy*, *Member*⟩ ⟨*Paper*, *hasTitle*, *Title*⟩	⟨*Author*, *author*, ≡⟩
20	⟨*Alice*, testify, *Paper*, *article*, ⟨*Paper*, *hasAuthor*, *Author*⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩ ⟨*Paper*, *wasReviewedBy*, *Member*⟩ ⟨*Paper*, *hasTitle*, *Title*⟩ ⟨*Paper*, *hasAuthor*, *Author*⟩	⟨*Author*, *author*, ≡⟩
21	⟨*Bob*, assert, *Paper*, *article*, {(⟨*Paper*, *wasReviewedBy*, *Member*⟩, ⟨*article*, *reviewer*, *pcmember*⟩) (⟨*Paper*, *hasTitle*, *Title*⟩, ⟨*article*, *entitled*, *title*⟩)) (⟨*Paper*, *hasAuthor*, *Author*⟩, ⟨*article*, *authoredby*, *author*⟩)} ⟨*Paper*, *article*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩ ⟨*article*, *authoredby*, *author*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩ ⟨*Paper*, *wasReviewedBy*, *Member*⟩ ⟨*Paper*, *hasTitle*, *Title*⟩ ⟨*Paper*, *hasAuthor*, *Author*⟩	⟨*Author*, *author*, ≡⟩
22	⟨*Alice*, accept, *Paper*, *article*, {(⟨*Paper*, *wasReviewedBy*, *Member*⟩, ⟨*article*, *reviewer*, *pcmember*⟩) (⟨*Paper*, *hasTitle*, *Title*⟩, ⟨*article*, *entitled*, *title*⟩)) (⟨*Paper*, *hasAuthor*, *Author*⟩, ⟨*article*, *authoredby*, *author*⟩)} ⟨*Paper*, *article*, ≡⟩	⟨*author*, *family*, *familyname*⟩ ⟨*author*, *named*, *firstname*⟩ ⟨*author*, *affiliated*, *researchlab*⟩ ⟨*article*, *reviewer*, *pcmember*⟩ ⟨*article*, *entitled*, *title*⟩ ⟨*article*, *authoredby*, *author*⟩	⟨*Author*, *hasSurname*, *Surname*⟩ ⟨*Author*, *hasForename*, *Name*⟩ ⟨*Author*, *affiliatedTo*, *University*⟩ ⟨*Paper*, *wasReviewedBy*, *Member*⟩ ⟨*Paper*, *hasTitle*, *Title*⟩ ⟨*Paper*, *hasAuthor*, *Author*⟩	⟨*Author*, *author*, ≡⟩ ⟨*Paper*, *article*, ≡⟩
23	⟨*Alice*, end, nil, nil, ⟩	-	-	-

TABLE 6.5: Continuation of the update of *Alice* and *Bob* private and public knowledge stores throughout the example dialogue run, from move 11-23.

Chapter 7

Strategic Decision Making: Metrics and Ranking Function

Chapter Outline

'Nothing in life is to be feared, it is only to be understood.' - Marie Curie

This thesis has introduced a novel selected sharing approach for ontology alignment, up to now there have been many assumptions made concerning the metrics used within the protocol. This chapter will define and formalise these metrics, and detail the parameters of the dialogue which will then be implemented in an evaluation of the approach in Chaper 8.

7.1 Dialogue Parameters

This section formalises the metrics used within the DbMN approach which in the previous chapters have been introduced yet assumed for the walkthrough of the dialogue. The metrics formalised in this chapter have been selected in order to empirically evaluate the dialogue approach using real world ontologies. The evaluation of the DbMN approach is presented in Chapter 8.

As dialogue presented in Chapter 6, allows two agents (each assuming their respective ontologies) to negotiate over candidate mappings by incrementally sharing select fragments of their knowledge bases. This approach is developed over a protocol and an agent strategy, which are formalised in Chapter 5, where assumptions were made over the choice of metrics used by the agents. This choice determined the strategy adopted by the agents, which affects the choice of moves made at each point in the dialogue.

To simplify the evaluation, these agents adopt the same strategies using the metrics described in this chapter. This eliminated the effect of heterogeneous metrics in the evaluation.

7.1.1 Dialogue metrics

The DbMN approach uses four metrics to support the dialogue. The first is a lexical similarity metric σ_l used by the opponent agent during the *propose* phase. This metric evaluates the string similarity in order to identify a candidate concept that is then proposed for the mapping. The agents also utilise the same ranking function rank throughout the *confirm* phase of the dialogue, which generates an ordering for the agents to share their knowledge fragments. The structural similarity metric is σ_s is used by both agents within the *propose* phase to determine the similarity of a pair of triples. The agents also use a neighbourhood σ_n metric in the *confirm* phase of the dialogue, which is used to assess the similarity of two neighbourhoods to ascertain if there is sufficient support for the proposition of an assertion. This allows the agents to either accept or reject a proposed similarity value supporting the current argument posed.

Lexical Similarity Metric

The function $\sigma_l : \mathsf{N_C} \times \mathsf{N_C} \rightarrow [0, 1]$ which returns the lexical similarity between the labels of two entity names $e, e' \in \mathsf{N_C}$, such that $\sigma_l(e, e') = 1$ iff $e = e'$ and 0 if the two labels are different.

The dialogue assumes that agents utilise a lexical string similarity metric in order to establish an 'anchor' concept which is then used to negotiate over throughout the dialogue. Cheatham [24] evaluated string similarity metrics and grouped them into three categories: *global* or *local*, *perfect sequence* or *imperfect sequence* and *set* or *word based*. The global and local category relates to the amount of information required by the metric to generate two strings as similar or not. Global metrics requires information related to all of the strings in one or both of the ontologies before it can attempt a match.

Local metrics require only the pair of strings which are attempting to be classified as similar. This local factor is important concerning. the DbMN approach due to the private nature of the dialogue in regards to sharing knowledge, which is why local based methods are preferable.

The perfect or imperfect sequence category relates to the ordering of characters in a string to be consider as a match. A perfect sequence requires the characters in the two strings to be in the same order for a match, whereas in an imperfect sequence a match can be still found if the characters are in a different order but amount of difference between the strings is under a given threshold.

The set or word based category relates to the degree of word overlap between the words within the two strings. Word based metrics perform well on long strings.

Using these categories taken from [24] the Jaro-Winker metric is classified as a non-set, local and imperfect sequence metric. These attributes of the metric and its consistently high performance over the evaluation conducted by Cheatham, make it an appropriate lexical similarity metric to be used within the DbMN approach.

String Metric	**Global/Local**	**Perfect/Imperfect**	**Set/Word based**
Exact	Local	Perfect-sequence	Non-set
Jaccard	Local	Perfect-sequence	Set
Jaro-Winkler	Local	Imperfect-sequence	Non-set
LCS	Local	Perfect-sequence	Non-set
Levenstein	Local	Imperfect-sequence	Non-set
Monge-Elkin	Local	Imperfect-sequence	Non-set
N-gram	Local	Imperfect-sequence	Non-set
Soft-Jacard	Local	Imperfect-sequence	Set
Soft TF-IDF	Global	Imperfect-sequence	Set
Stoilos	Local	Imperfect-sequence	Non-set
TF-IDF	Global	Perfect-sequence	Set

Table 7.1: String similarity metrics as categorised by Cheatham [24]

The Jaro-Winkler metric shares the same categorical attributes as N-gram, Monge-Elkin, Levenstein, and Stoilos metrics. For the precision value evaluating the conference datasets taken from the OAEI the Jaro-Winkler metric performed better than the Stoilos, Levenstein and Monge-Elkin metrics, and the same as the Ngram metric. For the recall over the same datasets the Jaro-Winkler metric performed better than the Monge-Elkin metric, and the same as the Ngram metric and Stoilos. For these performances over the precision and recall for the conference datasets, this metric was selected to be used with the DbMN approach.

The Jaro-Winkler lexical similarity metric, measures the string similarity based on the number and order of the common characters between two strings, and emphasises similarity which two strings have similar prefixes. This metric is defined [85] as:

$$d_w(s_1, s_2) = d_j(s_1, s_2) + l \cdot p \cdot [1 - d_j(s_1, s_2)]$$

where l is the length of the longest common prefix between the two strings s_1 and s_2, and p is a constant scaling factor that also controls the what emphasis placed on the similarity of the string prefixes.

Although the ontologies used within this study are modelled differently and likely heterogeneous in content they all model knowledge within a related domain (i.e. the conference organisation domain) and thus there is a degree of similarity in their use of terminology. Thus similarity metrics such as the Jaro-Winkler metric are better suited for such scenarios.

Further experimentation could be explored utilising various combinations of lexical similarity metrics and are discussed as future work in Chapter 8.

Acceptance of a lexical matching is determined by whether the lexical similarity of two terms exceeds the lexical threshold ϵ_l. This occurs at state S2 of the dialogue. At state S2 the opponent agent utilises the similarity metric in order to find a target entity that is most lexically similar to that posed by the proponent agent within the *initiate* move. This threshold establishes a minimum value that the metric must meet before an entity can be proposed in a *propose* move. The ϵ_l was initially set to a value of $\epsilon_l = 0$, in the evaluation of the DbMN approach, in order to isolate the influence of the ϵ_n value on the alignment generation. This value is set to 0 in order to test the performance of the approach using the neighbourhood similarity threshold, where it will not be influenced by the lexical similarity. This will mean that if the opponent agent finds an entity in its ontology with a lexical similarity to the entity posed in the initiate move with a low similarity value, it will still be proposed. This allows the implementation of the approach to be centred identifying heterogeneous concepts though investigating the neighbourhood rather than depending on lexical string similarity.

Neighbourhood Similarity metric

The function $\sigma_n : \{(\varpi, \varpi') \in \Pi \times \Pi \mid \varpi \in \Gamma, \varpi' \in O\} \rightarrow [0, 1]$ returns an aggregate similarity obtained by calculating all possible structural similarities between the triples in an agent's gamma store Γ and the triples in the disclosable fragment of the opponent's ontology O', such that $\sigma_n(\varpi, \varpi') = 1$ if the neighbourhood is structurally equivalent, and 0 otherwise. The neighbourhood similarity for a candidate mapping must equal or exceed a threshold in order to be accepted. This threshold has been kept the same for both agents, where the value is incremented for each of the three DbMN variants by a value of 0.025 from [0..1]. The rationale for this threshold value, is that when evaluating the alignments found by the approach, the precision and recall with respect to a reference alignment will converge on an accepted value where the alignment is most accurate in terms of the number of mappings found, and as the value gets closer to 1 the approach will overly filter the alignments meaning that no mappings will be accepted.

In the evaluation of the DbMN approach the threshold for the neighbourhood ϵ_n was initially incremented using a value of 0.1. In an initial run of the dialogue approach, this value showed results outlining a significant change between the results where $\epsilon_n = 0.4$ and $\epsilon_n = 0.7$ in order to further investigate this significant shift in values, the threshold ϵ_n was then incremented by a value of 0.025 to better detail the results.

Ranking Value

The rank function, $\mathsf{rank} : \mathsf{N_C} \to \mathcal{R} \subseteq \Pi$ returns an ordered list of triples ϖ in a path starting at some entity $e \in \mathsf{N_C}$, where $\forall \varpi_i, \varpi_j \in \mathcal{R} : \varpi_i \succcurlyeq \varpi_j$.

In order to define which triple is to be shared at each given *testify* move the ranking value is calculated over four dependant criteria: *Subsumption*, *Rarity*, *Connectivity* and *Popularity* such that the total weight assigned had a total of 1. Each of these criteria has an associated weight in the range of [0..1] which are used to determine the significance of each criteria and are formalised as $\langle w_s r_s, w_r r_r, w_c r_c, w_p r_p \rangle$ where:

Weight	**Rank score**
w_s = the weight for subsumption	r_s = the rank score for subsumption;
w_r = the weight for rarity;	r_r = the rank score for rarity;
w_c = the weight for connectivity;	r_c = the rank score for connectivity;
w_p = the weight for popularity;	r_p = the rank score for popularity.

1. **Subsumption**:

   This prioritises the concepts further down the subsumption hierarchy over those that are higher up. This is based on the principal of generality, that assumes that a more specific concept which appears further down a hierarchy can be a better classifier when determining a subject area [3]. For example, the top of a given hierarchy for a concept *Clothes* could be labeled as *Thing* and a concept lower in the hierarchy could be *Shirt*. Thus in order to establish meaning of the neighbourhood, without revealing all of the associated concepts, it would be more succinct to share the more specific *Shirt* concept rather than more general *Thing*. This Subsumption value is calculated using the following formula:

   $$Sub_i = \frac{H_c}{H_{depth}}$$

   where H_c is the depth, measured by the number of edges e of the concept i from the top concept $\top \in \delta$, and H_{depth} is the total depth of edges e in the knowledge base $\delta = (v, e)$ from the $\top \in \delta$ to the $\bot \in \delta$.

2. **Rarity**: The second criteria to be calculated is the rarity of the concept, weighting uniquely labeled concepts more highly than others. This is based on [3], however here the concepts are given a binary value, 0 if the label for a concept i appears more than once in the knowledge base δ, or a value of 1 if it is unique:

   $$Rar_i = 1 \text{ or } Rar_i = 0$$

3. **Connectivity**: The third criteria is the connectivity of a concept. This is found using the outgoing edges leaving from the concept $\delta^-(v)$. The basis of this value is taken from the notion of *connectivity* from [23] where it is stated that a concept is more central to an knowledge base if it has more outgoing edges. The idea of

measuring the incoming and outgoing edges of a concept to aid ranking is also seen in [3]; however here it is presented as a singular value, calculating the *popularity* of a concept by the edges regardless of their direction. In this work, this connectivity value is separated into two criteria popularity (incoming edges) and connectivity (outgoing edges), to take into account the direction of an edge and investigate the affect on the ranking.

This connectivity is found over a graph δ, which comprises a set of vertices V and edges E, such that $e \in E$ and $v \in V$. This value is calculated by dividing 1 by the total edges of a graph δ, $|E|$ then multiplying this to the number of the outgoing edges δ^- of a concept v^i:

$$Con_i = \tfrac{1}{|E|} \times \delta^-(i)$$

4. **Popularity**: The fourth criteria, popularity, is similar to the connectivity value; however measures the incoming edges of a concept.

   This popularity is calculated by dividing the total edges $|E|$ in the knowledge base δ, then multiplying this to the number of the incoming edges δ^+ of a concept v^i.

$$Pop_i = \tfrac{1}{|E|} \times \delta^+(v^i)$$

Each of these criteria are determined resulting in a normalised score in the range [0..1]. A final rank is then generated using a weighted average (mean) over the four criteria. It could be argued that connectivity and popularity should hold a higher weight over the others, as it is a feature that appears in alternative ontology rankings, such as CARRank [136] and DWRank [23].

The Structural Similarity Metric

This is the function $\sigma_s : \Pi \times \Pi \to [0, 1]$ that returns the structural similarity between two triples $\varpi, \varpi' \in \Pi$, such that $\sigma_s(\varpi, \varpi') = 1$ if the two triples are considered as equivalent, and 0 otherwise. This is calculated by finding an aggregated lexical similarity value using the Jaro-Winkler similarity metric [134], over the subject s, predicate p and object o in ϖ.

The Jaro-Winkler method previously discussed in 7.1.1 was selected, as using the Cheatham classification it is most suitable for non-set, local and imperfect metrics [24] as is the case in the DbMN approach. This value is then calculated against the aggregate similarity value for ϖ'.

At state S2 in Figure 5.4 the opponent agent may find one or more concepts that have equal lexical similarity to the entity proposed, however these concepts may have different meanings, indicating polysemous terms. In order for meaning to be established, the meanings of the polysemous terms needs to be explored within the neighbourhood. The moves between states S3 and S4 allow the agents to explore the nested hierarchy or subsumption of concepts, through the prioritisation of the 'is_a' relation. Across all the variants of the approach, this relation is weighted higher than all those coming from a concept, therefore in a *justify* move the 'is_a' relation will be shared first, followed by those in order established in the ranking from the above criteria.

7.2 Summary

This chapter details the methodology used for the evaluation implementation of the DbMN approach. This includes the parameters to which the experiments have been designed. This chapter has detailed the metrics, previously assumed, which are used within the DbMN approach. The metrics formalised in this chapter, are used within the evaluations of the dialogue approach presented in Chapter 8. The variants have been empirically evaluated using precision and recall compared to a benchmark standard, and finally this chapter has presented an overview of the DbMN performance in comparison to selected current systems.

In Summary of this Chapter:

- Formalised the criteria used for the ranking value.
- Formalised the metrics used by the agents within the dialogue approach.

Chapter 8

Evaluation

Chapter Outline

'The most damaging phrase in the language is: 'It's always been done that way'.' - Grace Hopper

Up to now this thesis has contextualised the elements of ontologies and dialogue into a dialogue approach for generating an alignment. Using this novel selective sharing approach agents can utilise the protocol in order to generate ontology alignments supported by semantic meaning, without sharing their full knowledge base.

The aim of this chapter is to provide evidence to support the DbMN and illustrating how the approach can successfully generate a meaningful alignment between two agents, without sharing the full ontologies a priori. Having presented a walkthrough example of the DbMN in Chapter 6 this chapter presents the implementation of three varied DbMN approaches (DbMN_5, DbMN_6 and DbMN_7) using real world ontologies, and empirically evaluates their performance in generating these alignments. These three variants of the approach are then evaluated in comparison to each other and finally in Section 8.4 the approaches are compared to current ontology matching systems. This chapter concludes with the overall findings of the DbMN approach, and discusses further experimentation for evaluating the dialogue approach.

8.1 DbMN Experiment Preliminaries

The implementation of the DbMN was designed using the protocol and formalisms detailed in Chapter 5 and Chapter 7 where two agents participate in a dialogue protocol in order to generate a meaningful alignment between their two knowledge bases. This implementation and the experiments designed, sets out to prove the overall hypothesis introduced in Chapter 5:

Is it possible to generate a meaningful alignment between two ontologies, without sharing both ontologies in full a priori?

This main research question has been used as the central focus of the design of the experiments and can be divided into additional hypothesis:

- Does the ordering of the proponent's signature affect the resulting alignment?
- How does the varying neighbourhood threshold influence the resulting alignment?
- How does using the proposed ranking method, over an alternative ordering affect the alignment generated?

These variables were used in order to investigate how the three variants of the DbMN approach, influenced the resulting alignments potentially in order to provide an optimum bound, or combination of variables where the approach performed most accurately over the given ontologies.

The following chapter details the experimental methodology and the parameters across all three variants including the variables and the strategies the agents utilise. This evaluation is conducted over the three variants of the approach DbMN_5, DbMN_6 and DbMN_7 and this chapter presents how the variants differ in strategy, variables and restrictions used. The variants are then compared using the generated alignments found by this approach and are also compared to the currently alignment systems detailed in Section 8.1.1

8.1.1 Hypothesis of experimentation

These experimental hypotheses were used in order to evaluate the results of the approach, in terms of precision and recall to a benchmark designed for this implementation detailed later in this section. The main hypothesis that this implementation aims to prove is:

Q1. *Can a plausible meaningful alignment be found, that compares to a benchmark standard, when two agents engage in a dialogical approach to ontology matching, rather than sharing their knowledge bases in full, a priori?*

This question is answered in its broadest sense, if the dialogue can produce a single mapping from one ontology to another, using the DbMN approach. However, in order to further investigate the results of the approach and analyse the effects of the agent strategies under different settings, the following subquestions were investigated to address the additional hypotheses:

Q2. *Can a plausible alignment be generated and maintain a level of privacy?*

Q3. *What are the influences of the signature order on the accuracy and correctness of the alignment?*

Q4. *What are the influences of the threshold levels on the accuracy and correctness of the alignment?*

Q5. *What are the influences of the neighbourhood sharing order on the alignment?*

It is important to note that the dialogue protocol approach is not analysed in terms of running time and complexity. The implementation of the protocol runs the dialogue, as a full iterative game over the proponent's signature until all the candidate concepts within this signature have been explored. This implementation results in a longer run time for large ontologies, and it has been assumed that the full set of the classes in the ontology of the proponent agent is in this signature. However, running time was not a focal point for this research and reducing this was not part of the experimentation.

As detailed in Chapter 5 there are three variants with adjusted strategies used to implement this DbMN approach. This section will now detail these three individual variants and present the results found through the experimentation conducted to evaluate the approach.

8.1.2 Evaluation methods

The evaluation of the approach is conducted over the precision and recall of the generated alignment A compared to a benchmark standard alignment R (detailed later in this section). The precision and recall of an alignment is a popular approach in evaluating ontology matching approaches.

Precision measures the level of correctness of the alignment generated in comparison to the reference alignment using the correct mappings found (those featuring in the benchmark) over the total number of mappings in the alignment. Precision is defined in [38] as a function $P : \Lambda \times \Lambda \rightarrow [0..1]$, where in this case Λ refers to an alignment such that R is the reference alignment and the precision of an alignment A is determined. Therefore:

$$P(A, R) = \frac{|R \cap A|}{|A|}$$

Recall measures the level of completeness of the alignment generated in comparison to the reference alignment. This recall uses the number of correct mappings found over the total number of expected mappings in the reference alignment. Recall is defined in [38] as a function $R : \Lambda \times \Lambda \rightarrow [0..1]$, where Λ refers to an alignment such that R is the reference alignment and the recall of an alignment A is determined. Therefore:

$$R(A,R) = \frac{|R \cap A|}{|R|}$$

The F-measure is the third value used to evaluate alignments generated by an ontology matching approach, which represents the harmonic mean of the precision and recall values presented above. Within this work the f-measure has been used and is presented as an averaged value over all the dataset pairs. These results are presented in Appendix A. The f-measure allows the results to be compared by the precision and recall where the f-measure is maximal. The f-measure where $\alpha = 0.5$ is defined in [38] as:

$$M_{0.5}(A,R) = \frac{2 \times P(A,R) \times R(A,R)}{P(A,R) + R(A,R)}$$

Precision, recall and the related f-measure between these values, have been calculated for each alignment generated over the incremented neighbourhood threshold by each of the three variants. These three measures have been selected as they are widely used, and accepted in evaluating current ontology matching systems [39], in terms of the accuracy of the mappings found in an alignment, over the total mappings found by the approach. The three DbMN variants are also evaluated in terms of the number of mappings each increment of the neighbourhood threshold finds$O$ that features in the corroborating *platinum standard* benchmark for the O to O' ontologies.

8.1.3 Datasets

The experiments for the three working versions of the dialogue protocol presented were conducted over all the paired datasets from the 2014/2015 from the OAEI *Ontology Alignment Evaluation Initiative* competition (see Table 8.2). The OAEI [67] provides access to real world ontologies, representing various domains. For the purposes of this work a series of ontologies from the *Conference Track* was used, which includes a reference or gold standard set of alignments for each pair.

This track consists of 16 real world ontologies[1] documenting the domain of 'organising an academic conference', all of which are heterogenous in their design. The OAEI conference dataset has been selected due to its extensive use in current ontology matching systems, providing a comparison to the results found using the dialogical approach presented in this work. The eight ontologies detailed in Table 8.1 were selected and used in the evaluation of this dialogical approach.

These ontologies all represent academic ontologies, and are relatively small in size, in terms of the number of classes presented in the ontology, ranging from a class size of 104 (edas) to 36 (cmt), presented in Table 8.2. With a pre-existing alignment, it provided a 'gold standard' set of alignment results between all the ontology pairs alongside evaluations by current ontology alignment systems. The 'gold standard' alignment is a

[1] It is important to note, that some of the ontologies, e.g. 'crs' were not included in this implementation due to reduced size, and 'iasted' was omitted due to errors found within the documentation. Other ontologies were not included as reference alignments could not be found on the OAEI

Table 8.1: Ontology Pairs

Proponent	**Opponent**	**Proponent**	**Opponent**
cmt	conference	conference	edas
cmt	confof	conference	sigkdd
cmt	sigkdd	conference	ekaw
cmt	ekaw	conference	confof
cmt	edas	edas	ekaw
confof	sigkdd	edas	sigkdd
confof	ekaw	ekaw	sigkdd
confof	edas	-	-

benchmark alignment generated by the OAEI and is wildly used as a reference alignment for evaluations [137]. They are manually developed by domain experts in order to provide a meaningful reference alignment for systems participating in the alignment initiatives.

Table 8.2: OAEI ontologies used in this implementation, included in conference track [67]

Name	**Number of Classes**	**Number of Object Properties**	**DL Expressivity**
cmt	36	49	ALCIN(D)
conference	60	46	ALCHIF(D)
confOf	39	13	SIN(D)
edas	104	30	ALCOIN(D)
ekaw	74	33	SHIN
sigkdd	49	17	ALEI(D)

By using the selected ontologies from the OAEI conference, the performance of the DbMN dialogical approach could be compared to that of the other alignment systems that participated in the 2014/15 Ontology Alignment Initiative [35], such as X-Map [32], AML [43], and LogMap [68], detailed in Chapter 3. A comparison with these approaches is discussed below and illustrated in Appendix A.

There are important differences however with the alignments generated by these other systems, as these other systems generate alignments between ontology pairs in an open environment when all the data is shared a priori. This is a fundamental difference compared to the dialogue based approach, however showing the alignment results generated by the dialogue based approach, provided at least a comparison independent of their mapping approach.

Utilising the reference alignments from the OAEI, the ontologies were assigned to the proponent and opponent agents respectively, generating the experimental dataset pairs in Table 8.1. The only assumptions made by the DbMN approach over the ontology language itself is that it conforms to the RDF/S axiomatic semantics and is representable as a graph. Thus although the different ontologies in Table 8.1 are expressed using different levels of expressivity (i.e. different flavours of description logics), their use is valid within the DbMN approach.

8.1.4 Benchmarks

The performance of the DbMN approach is evaluated using a benchmark derived from the reference alignment in the OAEI [67]. In current ontology alignment systems, a *Gold Standard* benchmark alignment is available for use when measuring the correctness and accuracy of the alignment generated by the approach. In this implementation of the DbMN approach, a derivative of this *gold standard* benchmark is used. The derivative proposed here as a *Platinum Standard* and is generated by pruning the *gold standard* reference alignment taken from the OAEI, to consider the most common mappings (i.e. those found by the majority of current alignment systems), and only class label correspondences. This pruning is utilised as the DbMN approach only looks for class label correspondences, therefore it is necessary to prune out those matches that include data or object type properties. Using the ontologies from the OAEI conference track; *cmt* as O and *conference* as O' the gold standard reference alignment is pruned into the platinum standard.

TABLE 8.3: Benchmark alignments for the ontologies O= cmt, O' = conference

Gold Standard		**Platinum Standard**	
O = **cmt**	O' = **conference**	O = **cmt**	O' =**conference**
Conference	Conference_volume	-	-
Preference	Review_preference	-	-
Author	Regular_author	-	-
Person	Person	Person	Person
email	has_an_email	-	-
Co-author	Contribution_co-author	Co-author	Contribution_co-author
PaperAbstract	Abstract	-	-
Document	Conference_document	Document	Conference_document
Review	Review	Review	Review
Conference	Conference	Conference	Conference
ProgramCommittee	Program_committee	ProgramCommittee	Program_committee
Chairman	Chair	-	-
SubjectArea	Topic	-	-
assignedByReviewer	invited_by	-	-
assignExternalReviewer	invites_co-reviewers	-	-
-	-	Reviewer	Reviewer
-	-	ProgramCommitteeMember	Program_committee
-	-	Paper	Paper

Table 8.3, illustrates the removal of the mappings found in the *gold standard*, to form the *platinum standard* including property mappings such as:

- $\langle assignedByReviewer \twoheadrightarrow invited\_by, \equiv\rangle$
- $\langle assignedExternalReviewer \twoheadrightarrow invites\_co\text{-}reviewers, \equiv\rangle$.

These are removed in the *platinum* benchmark, as only class labels are utilised in the signature to be mapped by the agents, and the properties will never be included. Other mappings are not included as they have not been found by the majority of systems; i.e. if the mapping is not found by the majority of current systems including: AML, AOT, AOTL, CROMatcher, DKPAOM, GMap, JarvisOM, Lily, MassMatch, Mamba, OMReasoner, RSDLWB and XMAP, the mapping will not be included in the *platinum standard*.

A comprehensive list of the number of entities, which are found in the both the OAEI *gold standard* benchmark, and the *platinum standard* presented for this thesis, for all the ontology pairs used in DbMN_5, can be seen in Table 8.4. The mappings in the both the OAEI *gold standard* benchmark and the *platinum standard* for the DbMN_6 and DbMN_7 for each of the individual ontologies can be found in Appendix B, and has not been included here for the sake of brevity.

TABLE 8.4: Number of Entities in the Benchmark Alignments taken from the OAEI Gold standard [67] and the Platinum standard generated for this implementation

Datasets		**Benchmarks**	
Ontology O	Ontology O'	*Gold Standard*	*Platinum Standard*
cmt	conference	15	9
cmt	confof	16	5
cmt	edas	13	8
cmt	ekaw	11	7
cmt	sigkdd	12	9
conference	confof	15	10
conference	edas	17	9
conference	ekaw	27	15
conference	sigkdd	15	11
confof	edas	22	10
confof	ekaw	20	16
confof	sigkdd	7	4
edas	ekaw	29	16
edas	sigkdd	15	7
ekaw	sigkdd	11	7

8.2 Experiment Parameters

In this section, the three sets of evaluations are described with details of the choice of parameters used. The results of each set of experiments are given in Section 8.3.

8.2.1 DbMN_5 Parameters

This section will detail the implementation and the empirical evaluation of the experiments run on the first of the three approaches, DbMN_5. This section will firstly introduce the dialogue parameters used that are specific to this variant, relating to the parameters defined in Chapter 7. These variable settings ([V.1..V.3]) are adjusted within the agents strategy and the task over which the agents negotiate an alignment. For these experiments, both agents are restricted to the same strategies, i.e. the metrics defined in this section hold for both agents participating in this dialogue.

V.1 **Lexical and Neighbourhood similarity threshold.** The use of the neighbourhood similarity threshold adjusted the strategy of the agent in the minimum bound of acceptance for the neighbourhood similarity metric as detailed in Chapter 7. The use of this threshold in the implementation of the DbMN_5, marks the first variable used in evaluating this approach. DbMN has been implemented, generating alignments over the two dataset ontologies, for each incremented value of the neighbourhood similarity threshold. In this implementation, the neighbourhood threshold is increased in increments of 0.025 from [0..1], producing 400 experiment files each generating an alignment for the two assigned ontologies. This allows the approach to be evaluated over these alignments when the neighbourhood similarity threshold is increased, meaning for a candidate mapping to be accepted, it has to be higher than this similarity threshold. The rationale behind this incremented neighbourhood threshold, was to investigate a range of this threshold where the alignment was found to be most accurate and correct (in terms of precision and recall), in comparison to a reference alignment. This was also investigated to find the range of the neighbourhood threshold where the alignments found the most number of *correct* mappings (i.e. those found which feature within the reference alignment). As detailed in Chapter 7 the lexical similarity threshold is set to 0 throughout the experiments, so that all possible candidate pairs were considered.

V.2 **Random vs Ordered signature.** The ordering of the signature is the next of the variables that was used to evaluate DbMN_5. In an initial run of DbMN_5, the experiment files were generated over an ordered signature. The choice of this ordering is arbitrary, and the entities were sorted alphabetically, comprising of all the class labels an ontology O. This created the order in which the proponent would initiate the concepts at state S1. The inclusion of the full classes in ontology in this signature was selected in order to test the approach when the full ontology was marked as disclosable, to maintain the reliability of the termination condition.

As this signature contains all the ontology classes, this results in the full disclosure of the proponents ontology O across the experiments. This however, allowed the opportunity for the opponent agent to maintain a level of privacy, and not share the full ontology O' throughout the dialogue process. Firstly the signature of the class labels was ordered alphabetically, and run alongside a randomly generated order of the signature, for each of the 0.025 increments of the neighbourhood similarity threshold. Over the initial run of experiments, this ordering of the signature had no effect on the alignments generated, and the results for both mapping the concepts in order or randomly generated the same candidate mappings and alignments outputted by the approach using balanced rank weights and the neighbourhood similarity threshold at 0. As this ordering was seen as a redundant variable for DbMN_5 and 6, it was discarded in the final experiments, however utilised in DbMN_7 as a result of the 1:1 mappings restriction, which would be influenced by the ordering of the signature $\boldsymbol{\Sigma}^t$. The results for only the randomly ordered signature have been evaluated, across the implemented versions DbMN_5 and 6, with the used of an ordered signature added to compare with a randomly ordered signature in DbMN_7.

V.3 **Ranking function**. The final variable used in the implementation of DbMN_5 was on the ranking function used by the agents, in determining which triple was shared, in a *testify* move throughout the dialogue exchange. This variable was set using the ranking function, generated over the aggregated four values, (*subsumption*, *rarity*, *connectivity* and *popularity*). As an alternative an alphabetic sharing was utilised, to test the influence of this ranking value on the outputted alignment results over the incremented neighbourhood similarity threshold values. The initial set of experiments on DbMN_5 showed no difference on the outputted alignments over each increment of the neighbourhood similarity threshold values, using either the ranking method weighted evenly or using the values of 0 or 1 weightings, to using an alphabetic sharing. This showed that the ranking had no influence on the generated alignments, and for the purposes of the evaluation of the results for DbMN_5 only the equally weighted ranking values have been used.

The variables detailed above were selected in order to evaluate the influences of these variables over the alignments generated by the approach across the incremented threshold levels, to find the optimum range where the approach performed best over its precision and recall values to that of the platinum standard. This would then provide an optimum range where the threshold found the most accurate alignments created.

The hypothesis of this decision was that as the neighbourhood threshold increased, the approach would find it harder to accept weaker and potentially *incorrect* mappings and the final alignment would reduce from the threshold value set from [0..1], becoming a more accurate alignment towards the value of 1. Alongside this hypothesis it was theorised that by increasing the neighbourhood threshold it would reduce the amount of the ontology O' shared throughout the dialogue.

It was also hypothesised that using a ranking method prioritising the triples to be shared using the four aggregated values, would produce better alignment results across these variables, by accepting more relevant mappings to the concept under negotiation. This however, was not proven to be true as none of the variants found a difference between the alignments generated using this ranking score. This could have been due to the weighting of the values in the ranking system using an aggregated value of 1 for the balanced weights. Setting these balanced weights all to a value of 1 may have produced a better representation of the ranking function, however this is presented as a potential avenue for further experiments.

8.2.2 DbMN_6 Paramaters

This section details the parameters for the experiments run on DbMN_6, including the evaluation metrics which were examined over a set of hypotheses designed to address the performance of the approach. DbMN_6 is the second variant of the approach and differs from DbMN_5, in the agent's strategy for the *testify* move. In comparison with DbMN_5, DbMN_6 shares the related concepts with a path length of 1, in a single move, rather than individually, therefore iterating through the *justify* and *testify* moves only once, as everything will be shared in a single iteration. The dialogue protocol for DbMN_6 is implemented over the following parameters:

V.1 **Lexical and neighbourhood similarity threshold**. The use of these thresholds in the evaluation of DbMN_6, is the same as DbMN_5, and generates 400 experiment files one for each of the 0.025 increments of this neighbourhood similarity threshold, each generating an alignment for the two assigned ontologies to the agent. The implementation of DbMN_6 uses the same restrictions on the lexical similarity threshold as DbMN_5, and for these experiments remains at 0 testing the neighbourhood similarity threshold independently.

V.2 **Random vs Ordered signature**. The ordering of the signature is the next of the variables that was used to evaluate DbMN_6. In an initial run of DbMN_6 similar to that of the DbMN_5 and similarly showed no difference between the random and ordered signature, so only a randomly generated order of the signature was used in the experiments for this approach.

V.3 **Ranking function**. The final variable used over in the implementation of DbMN_6 was on the ranking function used by the agents, determining which triple was shared, in a *testify* move throughout the dialogue exchange. In contrast to DbMN_5, DbMN_6 does not rank these triples, as they are all shared in a single *testify* move. This batch sharing approach was chosen as a variable in contrast to DbMN_5 in order to investigate if the same alignments were found, if the agents shared the neighbourhood of a concept in one iteration of the *testify-justify* move. This variable was used across the full incrementation from [0..1] of the neighbourhood similarity value.

Consider an example using the neighbourhood for the concept *Paper* taken from the *edas* ontology dataset. With this as ontology O' in DbMN_5 the neighbourhood is shared one triple at a time, until there are no more triples where *Paper* is the subject. DbMN_6 differs, such that the opponent agent shares the full neighbourhood of the following triples (where *Paper* is the subject) in one single testify move:

– $\langle Paper, is\_A, Document\rangle$
– $\langle Paper, isWrittenBy, Author\rangle$

This adjustment to the agent's strategy results in a singe iteration of the *testify-justify* loop for each of the concepts under negotiation, as all the triples are shared in a single move. As a consequence there are no further triples left to be shared in a second iteration of a *testify*. It was hypothesised that the alignments should remain the same, as the same bounds are being used, meaning that the mappings should be the same as those found in DbMN_5. However, the sharing will be different as this is adjusted for the agents in this version of the dialogue approach.

cmt-conference	confof-ekaw
AssociatedChair --> Abstract	Scholar --> OC_Chair
Rejection --> Rejected_contribution	Assistant --> Assigned_Paper
ProgramCommitteeMember --> Program_committee	Member_PC --> PC_Member
Acceptance --> Accepted_contribution	Camera_Ready_event --> Camera_Ready_Paper
AuthorNotReviewer --> Abstract	Registration_of_participants_event --> Regular_Paper
Meta-Reviewer --> Committee_member	Volunteer --> Document
Conference --> Conference	Administrative_event --> Academic_Institution
ProgramCommitteeChair --> Program_committee	Reception --> Location
Paper --> Paper	Country --> Contributed_Talk
Administrator --> Abstract	Author --> Abstract
Review --> Review	Reviewing_event --> Review
Document --> Committee	Tutorial --> Tutorial
Decision --> Person	Contribution --> Contributed_Talk
ProgramCommittee --> Program_committee	Trip --> Track
Author --> Abstract	Workshop --> Workshop
ConferenceChair --> Conference	Science_Worker --> Scientific_Event
SubjectArea --> Abstract	Person --> Person
ExternalReviewer --> Extended_abstract	Paper --> Paper
Meta-Review --> Conference_www	Working_event --> Workshop_Session
Bid --> Paid_applicant	Social_event --> Social_Event
Preference --> Presentation	Topic --> Tutorial
PaperFullVersion --> Paper	Company --> Conference
Person --> Person	Poster --> Poster_Paper
Reviewer --> Reviewer	Administrator --> Abstract
ConferenceMember --> Conference	Student --> Student
Chairman --> Chair	Conference --> Conference
User --> Poster	Banquet --> Student
Co-author --> Co-chair	Regular --> Regular_Paper
PaperAbstract --> Paper	Event --> Event
	Participant --> Paper
	Submission_event --> Submitted_Paper
	Short_paper --> Poster_Paper
	Member --> PC_Member
	Chair_PC --> PC_Chair
	Organization --> Organisation
	City --> Contributed_Talk
	Reviewing_results_event --> Review
	University --> University

FIGURE 8.1: 1:* restriction utilised in both DbMN_5 and DbMN_6, illustrated from the latter using dataset pairs O = cmt ➟ O'= conference and O= confof ➟ O'= ekaw.

8.2.3 DbMN_7 Parameters

This section details the parameters for the experiments run on the final variant of the dialogue based approach, DbMN_7. This section discusses the evaluation metrics which were examined and details the results of the variant over a set of hypotheses designed to address the performance of the process.

DbMN_7 has similarities to DbMN_5 in its use of the single neighbourhood sharing at a *testify* move, which will be detailed in variable *V.4* (below). However, in contrast to DbMN_5 and DbMN_6, DbMN_7 differs by introducing a one to one (i.e. injective) mapping restriction on the candidate mappings proposed. Using these variable settings DbMN_7 has been configured to investigate the differences of the alignments generated, in comparison to those in the previous variants, and finally in comparison to the current ontology matching systems.

One to one (i.e. injective) mappings within DbMN_7.

In the alignments illustrated in Figure 8.1, there are multiple instances where the mappings accepted into the alignments utilise an opponent concept from O' in more than one mapping. This can be seen in O = cmt ➟ O'= conference, with the concept *Abstract* $\in O'$ and also in O = confof ➟ O'= ekaw, with the concept *Review* $\in O'$:

$Abstract \in O' \langle AssociatedChair, Abstract, \equiv \rangle$

$$Abstract \in O'\langle AuthorNotReviewer, Abstract, \equiv\rangle$$
$$Abstract \in O'\langle Administrator, Abstract, \equiv\rangle$$
$$Abstract \in O'\langle Author, Abstract, \equiv\rangle$$
$$Abstract \in O'\langle SubjectArea, Abstract, \equiv\rangle$$
$$Review \in O'\langle Reviewing\_event, Review, \equiv\rangle$$
$$Review \in O'\langle Reviewing\_results\_event, Review, \equiv\rangle$$

This illustrates the properties of both DbMN_5 and DbMN_6 to permit one to many (i.e. 1:*) mappings which are accepted into an alignment by the participating agents. DbMN_7, seeks to investigate the dialogue approach in generating a mutually agreed alignment, when alignments are restricted to 1:1 mappings.

cmt-conference	confof-ekaw
ConferenceChair--> Conference_part	Volunteer--> Neutral_Review
Rejection--> Rejected_contribution	Assistant--> Assigned_Paper
AuthorNotReviewer--> Abstract	Scholar--> OC_Chair
ConferenceMember--> Conference_fees	Administrator--> Industrial_Paper
Meta-Reviewer--> Committee_member	Member_PC--> OC_Member
PaperFullVersion--> Paid_applicant	Camera_Ready_event--> Camera_Ready_Paper
Conference--> Conference	Registration_of_participants_event--> Regular_Paper
Acceptance--> Active_conference_participant	Administrative_event--> Academic_Institution
Paper--> Paper	Reception--> Location
PaperAbstract--> Camera_ready_contribution	Organization--> Organising_Agency
Review--> Review	Conference--> Conference_Paper
Document--> Committee	Author--> Abstract
AssociatedChair--> Accepted_contribution	Contribution--> Contributed_Talk
Person--> Publisher	Trip--> Track
ProgramCommittee--> Program_committee	Workshop--> Workshop
Decision--> Contribution_co-author	Person--> Person
ProgramCommitteeMember--> Steering_committee	Paper--> Paper_Author
Author--> Tutorial	Science_Worker--> Scientific_Event
ExternalReviewer--> Extended_abstract	Country--> Document
Meta-Review--> Conference_www	Working_event--> Workshop_Session
Preference--> Presentation	Social_event--> Social_Event
Administrator--> Important_dates	Topic--> Tutorial
SubjectArea--> Poster	Company--> Conference
Reviewer--> Reviewer	Reviewing_event--> Research_Institute
ProgramCommitteeChair--> Organizing_committee	Student--> Student
Chairman--> Chair	Tutorial--> Tutorial_Chair
Co-author--> Co-chair	Banquet--> Organisation
Bid--> Invited_speaker	Event--> Event
User--> Person	Participant--> Paper
	Submission_event--> Submitted_Paper
	Poster--> Poster_Session
	City--> Invited_Talk
	Short_paper--> Poster_Paper
	Member--> PC_Member
	Chair_PC--> PC_Chair
	Regular--> Regular_Session
	Reviewing_results_event--> Review
	University--> University

Figure 8.2: 1:1 restriction utilised in DbMN_7, illustrated using dataset pairs O = cmt ➟ O'= conference and O= confof ➟ O'= ekaw

Figure 8.2 illustrates the resulting 1:1 restriction on the dataset pairs O = cmt ➟ O'= conference and O= confof ➟ O'= ekaw. This shows that the mappings for the concepts can only use the concepts proposed from O' once in each alignment:

$$Abstract \in O'\langle AuthorNotReviewer, Abstract, \equiv\rangle$$
$$Review \in O'\langle Reviewing\_results\_event, Review, \equiv\rangle$$

This figure illustrates a 1:1 mapping restriction over the 1:* mappings seen in Figure 8.1 for the ontologies O = cmt ➟ O'= conference. With this restriction it can be seen that the concepts $Abstract \in O$ and$Review \in O'$ are only mapped once with DbMN_7. This 1:1 mapping restricts the opponent agent in a *propose* move, where they cannot reuse concepts that have previously been shared and successfully mapped, into an alignment. This utilise the trace in the commitment stores shared by the agents, in order to only propose new mappings for a candidate correspondence. DbMN_7 has been implemented over the following variables:

V.1 **Lexical and neighbourhood similarity threshold**. The use of these thresholds in the implementation of DbMN_7 is the same as the two previous variants detailed in this experimentation, and generates 400 experiment files, one for each of the 0.025 increments of this neighbourhood similarity threshold, each generating an alignment for the two assigned ontologies to the agent. This was done to investigate the most influential range of the neighbourhood similarity threshold, where DbMN_7 performs best in comparison to a reference alignment, in terms of the number of mappings it found which are categorised as *correct* mappings (those featuring in the reference alignment) and those which were deemed *incorrect*.

V.2 **Ranking function**. The final variable used over in the implementation of DbMN_7 was on the ranking function used by the agents, determining which triple was shared, in a *testify* move throughout the dialogue exchange. DbMN_7 uses the same function of sharing a single element of a neighbourhood in a *testify* move as detailed in DbMN_5. This single sharing was chosen as a variable in order to investigate the combination of the 1:1 mappings restricted in DbMN_7, with this use of multiple iteration of the *testify-justify* move, on the alignments generated by the variant, across the full incrementation from [0..1] of the neighbourhood similarity value.

V.3 **Random vs Ordered signature**. The ordering of the signature is the next of the variables that was used to evaluate DbMN_7 and was used due to the 1:1 mapping restriction. It was hypothesised, that as a result of the 1:1 mapping restricted on the candidate mappings, that the ordering of the alignment would have an effect on the mappings accepted, as concepts can only be proposed by the opponent agent, if they do not currently appear in a mapping for another concept. In order to investigate the ordering on the effects of the alignment, the random signature and an alphabetic sort on the signature was performed which had no effect on the alignment.

This ordering was investigated in comparison to the random signature results generated by DbMN_7, in terms of precision and recall of the final alignment across the incremented ϵ_n from [0..1]. The results were also evaluated in terms of the total number of mappings and the correct mappings found by DbMN_7 using an ordered signature in comparison to the random signature.

8.3 Empirical Evaluation

8.3.1 Empirical evaluation of approach; DbMN_5

Using the variables detailed above in the previous section, the dialogue protocol DbMN_5 was evaluated to examine the experimental questions detailed in the experiment preliminaries, and are evaluated with the corresponding related results found by DbMN_5.

[Q.1] *Can the dialogue approach find a plausible meaningful alignment, that compares to a benchmark standard, when two agents engage in a dialogical approach to ontology matching, rather than sharing their knowledge bases in full, a priori?*

The DbMN_5 variant was able to generate meaningful alignments containing *correct* mappings *(mappings found in the platinum standard)*, across the ϵ_n bounds of [0..0.650] for all the data set pairs used in this evaluation.

	SYSTEM:	N_ 0.500	N_ 0.525	N_ 0.550	N_ 0.575	N_ 0.600	N_ 0.625	N_ 0.650	N_ 0.675	N_ 0.700	N_ 0.725	N_ 0.750	N_ 0.775	N_ 0.800	N_ 0.825	N_ 0.850	N_ 0.875	N_ 0.900	N_ 0.925	N_ 0.95	N_ 0.975	N_ 1
cmt-confof	Total Maps	26	26	21	19	14	8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cmt-conference	Total Maps	24	24	16	16	15	8	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	4	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
cmt-edas	Total Maps	27	26	24	20	20	15	8	5	5	4	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	7	7	6	6	5	5	3	3	2	0	0	0	0	0	0	0	0	0	0	0
cmt-ekaw	Total Maps	26	25	24	21	21	17	10	9	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
cmt-sigkdd	Total Maps	25	20	17	16	16	15	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	5	4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
conference-confof	Total Maps	52	52	46	34	25	22	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	5	5	5	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
conference-edas	Total Maps	49	48	42	32	17	14	13	3	3	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	6	6	4	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-ekaw	Total Maps	55	55	54	46	35	22	15	8	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	11	11	10	10	9	6	5	2	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-sigkdd	Total Maps	38	33	22	20	10	6	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	5	5	4	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-edas	Total Maps	30	25	19	14	12	9	6	4	2	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	4	4	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-ekaw	Total Maps	34	34	29	23	17	13	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	9	9	9	8	7	6	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0
confof-sigkdd	Total Maps	28	28	22	17	15	6	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
edas-ekaw	Total Maps	87	86	83	78	67	50	38	27	18	8	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	12	11	11	11	9	6	4	4	3	2	0	0	0	0	0	0	0	0	0	0	0
edas-sigkdd	Total Maps	76	65	51	39	36	26	19	11	11	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	4	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
ekaw-sigkdd	Total Maps	61	56	54	36	23	18	16	5	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	5	4	4	4	4	4	3	3	3	0	0	0	0	0	0	0	0	0	0	0

FIGURE 8.3: Heatmap illustrating the upper bound of the ϵ_n, where the area in red indicates that no alignment can be found using DbMN_5.

Figure 8.3 illustrates that, from ϵ_n= 0.650 using the ontologies O= confof ➟ O'= sigkdd, the DbMN_5 variant can no longer find an alignment which includes mappings that are found in the benchmark. This Figure 8.3 also shows that the alignment at ϵ_n= 0.650 for these selected ontologies, consists of an alignment with 3 *correct* mappings. Following from ϵ_n= 0.650 the ontologies O= cmt ➟ O'= confof and O= conference ➟ O'= confof are both unable to find alignments from ϵ_n= 0.650 onwards using DbMN_5 and both at ϵ_n= 0.650 find no alignment. The remaining ontologies no longer find an alignment for ϵ_n= [0.725...0.750]. This illustrates that the bound for the neighbourhood threshold level of ϵ_n= 0.750 is too high for any mappings to be mutually accepted by the agents and therefore no alignments can be generated by DbNM_5.

These results prove the main hypothesis of this experimentation in that from the neighbourhood threshold values of [0..0.750] a meaningful alignment can be generated between the two agents using this dialogue protocol to incrementally share their knowledge bases, rather than full a priori. This shows that the top bound for the neighbourhood threshold value in order for this variant to work is 0.750.

[Q.2] *Can the dialogue approach generate a plausible alignment, and maintain a level of privacy?*

Q.1 investigates the fact that DbMN_5 can generate meaningful alignments: however over the ontologies presented in this implementation, there is a bound between ϵ_n=[0..0.750] where this holds. From ϵ_n=[0.750...1], no alignment can be found across any of the ontology pairs. From these results, *Q.2* addresses the number of *private* concepts from O' that are not disclosed throughout the iteration of the dialogue over the proponents signature. The variable setting, *V.3* addresses the fact that DbMN_5 includes all the concepts from ontology O in the signature to be mapped, therefore there is no privacy for the proponent agent using DbMN_5. There is however the opportunity for O' to explore the degree to which the concepts are disclosed. This is illustrated in the results presented in Table 8.5, where the unshared concepts of the ontology are detailed.

Table 8.5: Table presenting the level of privacy for the opponent agent's ontology O'

O -$O\prime$	$e \in O'$ shared at $\epsilon_n = 0$	$e \in O'$ shared at $\epsilon_n = 1$	total no. entities $e \in O'$	$e \in O'$ unshared between ϵ_n0-1	% of $O\prime$ shared at $\epsilon_n = 0$	% of $O\prime$ shared at $\epsilon_n = 1$
cmt-conference	26	25	60	1	43.33%	41.67%
cmt-confof	23	22	39	1	58.97%	56.41%
cmt-edas	33	31	104	2	31.73%	29.81%
cmt-ekaw	30	27	74	3	40.54%	36.49%
cmt-sigkdd	21	23	50	+2	42%	46%
conference-confof	31	30	39	1	79.48%	76.92%
conference-edas	50	42	104	8	48.08%	40.38%
conference-ekaw	50	45	74	5	67.57%	60.81%
conference-sigkdd	32	30	50	2	64%	60%
confof-edas	37	36	104	1	35.58%	34.62%
confof-ekaw	36	32	74	4	48.65%	43.24%
confof-sigkdd	24	26	50	+2	48.65%	52%
edas-ekaw	51	46	74	5	68.92%	62.16%
edas-sigkdd	39	35	50	4	78%	70%
ekaw-sigkdd	37	33	50	4	74%	66%

It can be seen in Table 8.5 that from the values of ϵ_n=[0..1] over for the following ontologies that there is a reduction in the number of concepts shared from ontology O':

O = cmt ➟ O' = conference
O = cmt ➟ O' = ekaw
O = cmt ➟ O' = confof
O = cmt ➟ O' = edas
O = confof ➟ O' = ekaw
O = confof ➟ O' = edas
O = conference ➟ O' = edas
O = conference ➟ O' = sigkdd
O = conference ➟ O' = ekaw
O = conference ➟ O' = confof
O = edas ➟ O' = ekaw
O = edas ➟ O' = sigkdd
O = ekaw ➟ O' = sigkdd

With these ontologies the number of concepts unshared from ϵ_n=[0..1] ranges from 1 concept to 8 concepts. There is however two cases where 2 more concept shared between ϵ_n=[0..1]. These two instances are:
O = cmt ➟ O' = sigkdd O = confof ➟ O' = sigkdd

This is a result of the increased neighbourhood threshold on the mapping acceptance. Table 8.6 illustrates that for the ontologies O=cmt ➟ O'=sigkdd, it can be seen that the concept *Deadline_Author_notification* is shared when the DbMN_5 threshold is set to 1, but not when this value is set to 0. This is to be expected as the agents require more triples to be shared when developing the premise (Pr) when the neighbourhood threshold is set higher.

	Neighbourhood threshold ϵ_n= 0	**Neighbourhood threshold ϵ_n= 1**
1	⟨Proponent, initiate, Co-author, nul, nul⟩	⟨Proponent, initiate, Co-author, nul, nul⟩
2	⟨Opponent, propose, Co-author, Author, nul⟩	⟨Opponent, propose, Co-author, Author, nul⟩
3	⟨Proponent, justify, Co-author, Author, nul⟩	⟨Proponent, justify, Co-author, Author, nul⟩
4	⟨Opponent, testify, Co-author, Author, ⟨Author, submit, Paper⟩⟩	⟨Opponent, testify, Co-author, Author ⟨Author, submit, Paper⟩⟩
5	⟨Proponent, assert, Co-author, Author, ⟨Author, submit, Paper⟩⟩	⟨Proponent, justify, Co-author, Author, nul⟩
6	⟨Opponent, accept, Co-author, Author, ⟨Author, submit, Paper⟩	⟨Opponent, testify, Co-author, Author, ⟨Author, Notification_until, Deadline_Author_notification ⟩
-	**Dialogue ended on concept 'Co-author'**	
7		⟨Proponent, justify, Co-author, Author, nul⟩
8		⟨Opponent, testify, Co-author, Author ⟨Author, obtain, Award⟩⟩
9		⟨Proponent, justify, Co-author, Author, nul⟩
10		⟨Opponent, testify, Co-author, Author ⟨Author, IS_A, Speaker⟩⟩
11		⟨Proponent, justify, Co-author, Author, nul⟩
12		⟨Opponent, reject, Co-author, Author
13		⟨Proponent, fail, Co-author, Author
-		**Dialogue ended on concept 'Co-author'**

TABLE 8.6: Difference in sharing over the ϵ_n=[0..1] values, using O=cmt ➟ O'=sigkdd

This difference in sharing is a result of the neighbourhood threshold set to ϵ_n= 0, where the proponent agent has enough support to accept the mapping without requesting a second triple for the premise. Therefore, this second triple containing this concept *Deadline_Author_notification* is not shared when the neighbourhood threshold is 0 This illustrates that the approach is working correctly, showing how the premise for the support of the mappings can differ with this increased neighbourhood threshold. Table 8.5 also illustrates the percentage of O' which is shared when using DbMN_5. Across all the ontologies, there is maximum of 79.48% of concepts in ontology O' and a minimum of 31.73% shared at $\epsilon_n = 0$, and a maximum of 76.92% of concepts in ontology O' and a minimum of 29.81% shared at $\epsilon_n = 1$.

An example of the concepts unshared in O' can be seen in Figure 8.4. This illustrates the concept in the ontology O' where this is the conference dataset, and cmt is the O dataset. In this figure the unshared concepts are represented as the smaller light grey

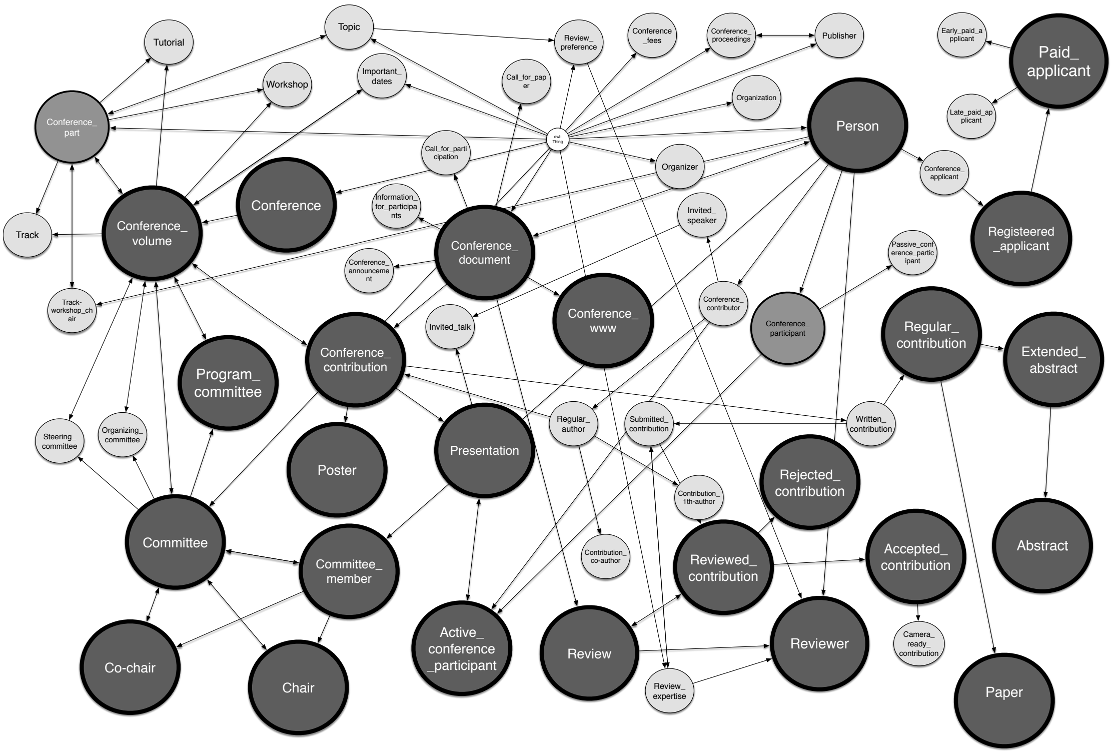


FIGURE 8.4: Opponent's shared concepts from their committed ontology. The darker and larger the node, the more the concept is shared, the lightest are the unshared concepts in agent's ontology.

nodes in the ontology as a directed graph. This illustrates the relation of shared concepts, to those concepts that are not disclosed throughout the process of the dialogue where the ϵ_n value is set to 1. Using these detailed settings for the experiments, the 25 concepts shared from O' *(shown in Figure 8.4 in the dark grey)* leave 35 concepts unshared *(shown in Figure 8.4 in the light grey)*.

Figure 8.4 shows there is no pattern in relation to sharing with the concepts position in the knowledge base in terms of hierarchy or connectivity. This suggests that bottom level concepts can be either shared (e.g. *Paper*) or kept private (e.g. *Important_dates*). This lack of correlation in the sharing of the concepts is due to the ranking value which is not presenting effect of the alignment, as the alignments sizes are the same if the ranking is used or if the ontologies are shared alphabetically.

Although these 25 concepts are shared this presents only 41.67% of O', where if using a centralised approach, the ontologies are shared in full a priori. This suggests that the DbMN approach can maintain a level of privacy and allow the opponent to share less of its ontology where in the worse case, no accepted mappings are found and no alignment is still generated.

Overall the results of this evaluation for this experimental question *Q.2* show a decrease in number of concepts shared by the opponent as the neighbourhood threshold increases. The proponent's signature consists of their full ontology, there is no unshared concepts from O in a run of DbMN_5, however there is never a case using DbMN_5, where the full O' is shared.

[Q.3] *What are the influences of the threshold levels on the accuracy and correctness of the alignment*

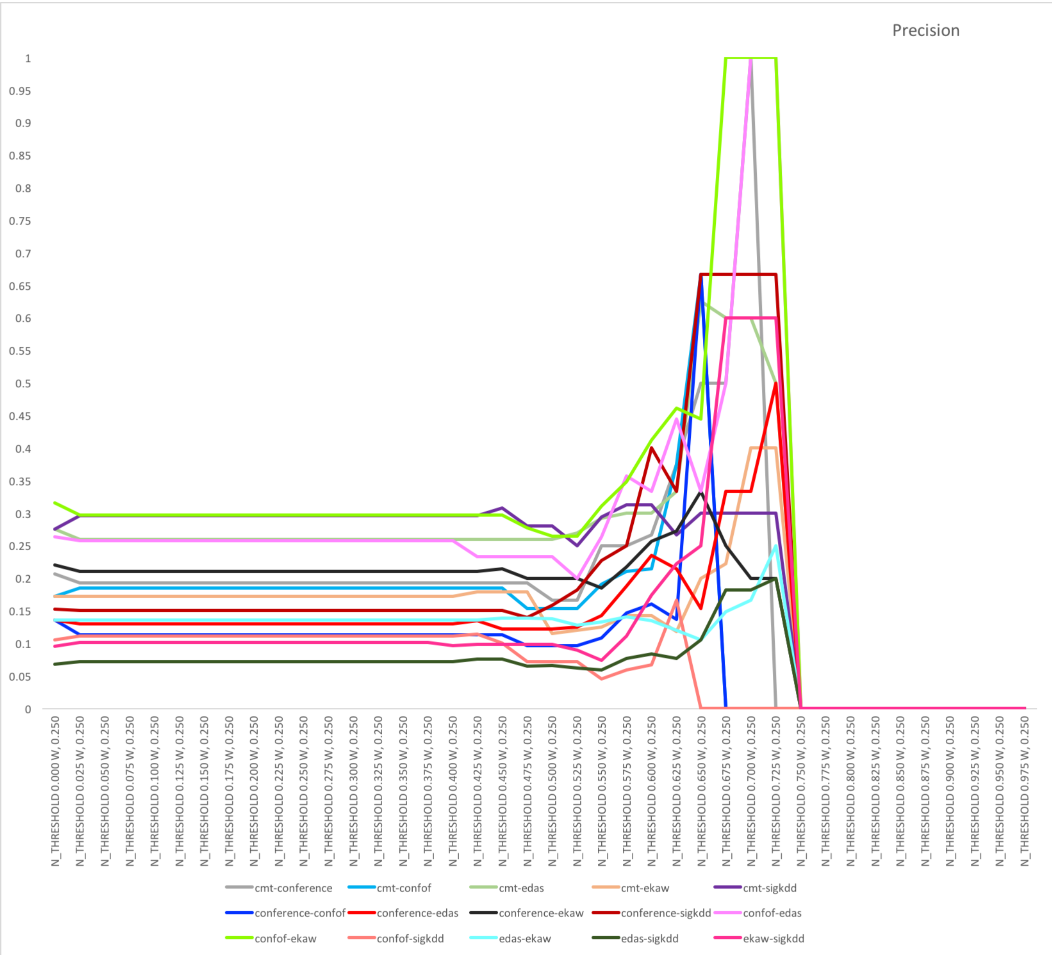


FIGURE 8.5: Precision curves for all the dataset pairs used in this experimentation.

The neighbourhood threshold as discussed in the preliminaries, has been tested in isolation to investigate the influence of locality in the generation of an alignment generation, over the use of just a lexical match. When this neighbourhood threshold is increased from [0..1] the overall results for all the dataset pairs present a negative correlation over the number of mappings accepted by the agents into the alignment. The results of the influence of this variable, have been divided over two evaluation metrics: firstly the precision of the alignment generated by the approach in relation to *correct* mappings found that appear in the benchmark; and secondly, the recall of the alignment generated with respect to the number of mappings found by DbMN_5 over the incremented neighbourhood threshold value.

Influence of the neighbourhood threshold value (ϵ_n) on the precision of the outputted alignment.

The Figure 8.5 shows the precision curves for all the dataset pairs used in this experimentation. The precision for DbMN_5, is generally a positive correlation with the increased neighbourhood threshold value. This can be seen at a value of 1 for the following ontologies:

O = confof ➟ O' = ekaw

O = confof ➟ O' = edas

O = cmt ➟ O' = conference

DbMN_5 fails to find an alignment at the value of ϵ_n= 0.750 however, from the ϵ_n= [0..0.750] this positive correlation proves the hypothesis that as the neighbourhood threshold is increased the approach finds more correct mappings, and at ϵ_n= 0.750 the DbMN_5 overfits the alignments, illustrating a top bound for the neighbourhood value to produce a meaningful alignment. Figure 8.5 illustrates a lower precision curve for the ontologies O= edas ➟ O'= sigkdd where the highest precision value is 0.2 at ϵ_n= 0.725, and the lowest value of 0.05 at ϵ_n= 0.550 with the precision tailing off to a value of 0 at ϵ_n= 0.750. This precision value for these ontologies is significantly lower than that of the previously discussed results. This could be a result of the O= edas, being the largest dataset used in this evaluation by at least 30 concepts, thus the O' has to map to a larger signature than the number of concepts it has in O', meaning that these mappings could potentially be weaker than if there was fewer concepts in the signature.

It can also be seen in Figure 8.5 that there are some case where the precision falls slightly before it rises. This is generally between the values of [0.525..0.650]. This is a result of the approach reducing the number of *incorrect* mappings it finds, however leaving the correct mappings unchanged.

Influence of the neighbourhood threshold value (ϵ_n) on the recall of the outputted alignment.

The Figure 8.6 illustrates the recall curves for all the dataset pairs used in this experimentation. The recall for DbMN_5 presents a negative correlation with the increased neighbourhood threshold value. This recall value has an upper value of 1 for the following ontologies:

O = ekaw ➟ O' = sigkdd

O = cmt ➟ O' = confof

O = confof ➟ O' = sigkdd

O = confof ➟ O' = edas

O = edas ➟ O' = sigkdd

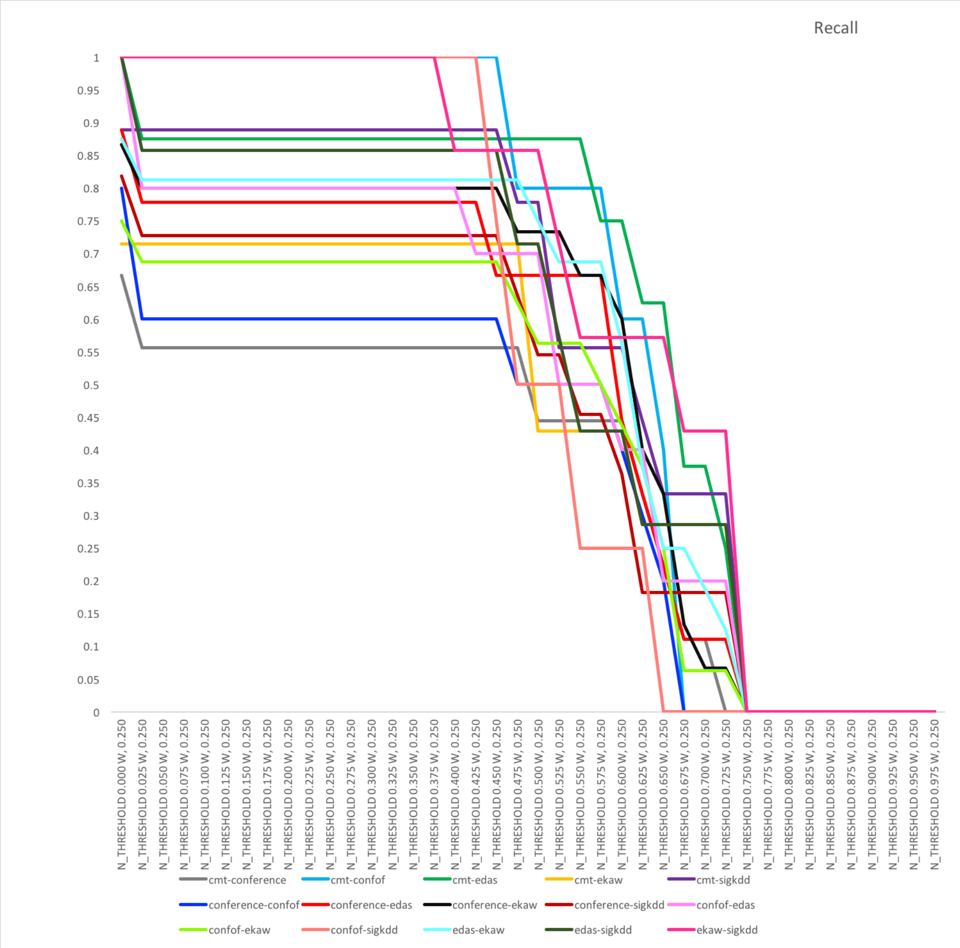


FIGURE 8.6: Recall curves for all the dataset pairs used in this experimentation.

The recall values decrease steadily across all the ontologies pairs from the neighbourhood values of [0..0.750] corresponding to the upper bound of this threshold in accordance with the precision curves. The negative correlation found between the recall and the incremented neighbourhood threshold values supports the hypothesis that as the neighbourhood threshold increases, the approach finds fewer mappings as it becomes harder for *incorrect* mappings to be accepted. As seen with the precision value the recall drops to 0 at ϵ_n= 0.750, where the approach overfits the alignments finding no acceptable mappings.

The intersection of the precision and recall curves can be seen in the graphs in Appendix A and is between the values of ϵ_n=[0.600..0.625] for all the ontologies. It is at this point therefore that DbMN_5 finds the best combination of results for the precision and recall. It is at ϵ_n=[0.600..0.625] that the f-measure is the highest, illustrated in Figure A.37. This illustrates the preferred neighbourhood setting for these dataset pairs where the corroboration between the most accurate alignment and the most complete alignment returns the best overall mappings.

[Q.4] *What are the influences of the neighbourhood sharing order on the alignment?*

From the experiments evaluating DbMN_5, no difference was established regarding the ranking and order of disclosing concepts either alphabetically of using the aggregated rank function. Figure 8.3 illustrates the total number of mappings found by DbMN_5 over the ontologies O = cmt ➟ O'= conference and the number of correct mappings found between the different methods of sharing. It can be seen that both methods find the same number of mappings and furthermore the two sets of mappings are equivalent.

Summary of DbMN_5.

To summarise, given DbMN_5 the dialogue approach provides a level of privacy for the opponent agent and also between the values of ϵ_n =[0..0.650] can generate a meaningful alignment between the two ontologies using the majority of dataset pairs used in this experimentation. These results for DbMN_5 supports the hypothesis that by using this dialogical approach, a meaningful alignment can be found whilst maintaining a selective sharing strategy for the opponent agent. Furthermore DbMN_5 supports the hypothesis that using this dialogue protocol a meaningful alignment can be found over two ontologies between the neighbourhood threshold values of [0..0.625].

This supports the issue of allowing an element of privacy for the opponent agent, in that they do not share their full ontology. The results demonstrate that, the approach still permits some degree of privacy when the neighbourhood acceptance threshold is 0. DbMN_5 has been adjusted to adapt the agents sharing strategy at the *testify* move to allow the speaking agent to share their full neighbourhood of the concept in a single move, rather than iterating through each element in the neighbourhood over successive moves.

8.3.2 Empirical evaluation of approach; DbMN_6

Using the evaluation settings described in Section 8.2, the dialogue protocol DbMN_6 was evaluated to examine the experimental questions *Q.1..Q.4* detailed in the experiment preliminaries. The results are then compared to those found with DbMN_5.

[Q.1] *Can the dialogue approach find a plausible meaningful alignment, that compares to a benchmark standard, when two agents engage in a dialogical approach to ontology matching, rather than sharing their knowledge bases in full, a priori?*

The DbMN_6 variant supports this hypothesis, and was able to generate meaningful alignments containing *correct* mappings across the ϵ_n bounds of [0..0.625] for all the data set pairs used in this evaluation. Figure 8.7 illustrates the gradual declining in both the number of mappings by DbMN_6, and the number of mappings found featuring in the platinum standard by DbMN_6. Figure 8.8 illustrates that from ϵ_n = 0.625 using the ontologies O= confof ➟ O'= sigkdd DbMN_6 can no longer find an alignment which includes mappings that are found in the benchmark.

	SYSTEM:	N_ 0.000	N_ 0.025	N_ 0.050	N_ 0.075	N_ 0.100	N_ 0.125	N_ 0.150	N_ 0.175	N_ 0.200	N_ 0.225	N_ 0.250	N_ 0.275	N_ 0.300	N_ 0.325	N_ 0.350	N_ 0.375	N_ 0.400	N_ 0.425	N_ 0.450	N_ 0.475	N_ 0.500
cmt-confof	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
cmt-conference	Total Maps	29	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	24
	Correct Maps	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4
cmt-edas	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
cmt-ekaw	Total Maps	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	28	28	28	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3
cmt-sigkdd	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	25	25
	Correct Maps	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7
conference-confof	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	52	52
	Correct Maps	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
conference-edas	Total Maps	59	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	52	49	49	49
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6
conference-ekaw	Total Maps	59	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	56	55	55
	Correct Maps	13	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	11
conference-sigkdd	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	50	38
	Correct Maps	9	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	6
confof-edas	Total Maps	38	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	30	30	30	30
	Correct Maps	10	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7	7	7
confof-ekaw	Total Maps	38	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	36	34
	Correct Maps	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	10	9
confof-sigkdd	Total Maps	38	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	35	30	28	28
	Correct Maps	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	2
edas-ekaw	Total Maps	103	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	94	94	87
	Correct Maps	14	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	12
edas-sigkdd	Total Maps	103	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	79	79	77	76
	Correct Maps	7	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
ekaw-sigkdd	Total Maps	73	69	69	69	69	69	69	69	69	69	69	69	69	69	69	69	62	61	61	61	61
	Correct Maps	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6	6	6

FIGURE 8.7: Total mappings and correct mappings from $\sigma_n = [0.0..0.5]$ using DbMN_6, where red depicts a lower number of mappings found, and green a higher number of mappings found.

	SYSTEM:	N_ 0.500	N_ 0.525	N_ 0.550	N_ 0.575	N_ 0.600	N_ 0.625	N_ 0.650	N_ 0.675	N_ 0.700	N_ 0.725	N_ 0.750	N_ 0.775	N_ 0.800	N_ 0.825	N_ 0.850	N_ 0.875	N_ 0.900	N_ 0.925	N_ 0.95	N_ 0.975	N_ 1
cmt-confof	Total Maps	26	26	21	19	14	8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cmt-conference	Total Maps	24	24	16	16	15	8	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	4	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
cmt-edas	Total Maps	27	26	24	20	20	15	8	5	5	4	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	7	7	6	6	5	5	3	3	2	0	0	0	0	0	0	0	0	0	0	0
cmt-ekaw	Total Maps	26	25	24	21	21	17	10	9	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
cmt-sigkdd	Total Maps	25	20	17	16	16	15	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	5	4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
conference-confof	Total Maps	52	52	46	34	25	22	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	5	5	5	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
conference-edas	Total Maps	49	48	42	32	17	14	13	3	3	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	6	6	4	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-ekaw	Total Maps	55	55	54	46	35	22	15	8	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	11	11	10	10	9	6	5	2	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-sigkdd	Total Maps	38	33	22	20	10	6	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	5	5	4	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-edas	Total Maps	30	25	19	14	12	9	6	4	2	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	4	4	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-ekaw	Total Maps	34	34	29	23	17	13	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	9	9	9	8	7	6	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0
confof-sigkdd	Total Maps	28	28	22	17	15	6	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
edas-ekaw	Total Maps	87	86	83	78	67	50	38	27	18	8	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	12	11	11	11	9	6	4	4	3	2	0	0	0	0	0	0	0	0	0	0	0
edas-sigkdd	Total Maps	76	65	51	39	36	26	19	11	11	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	4	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
ekaw-sigkdd	Total Maps	61	56	54	36	23	18	16	5	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	5	4	4	4	4	4	3	3	3	0	0	0	0	0	0	0	0	0	0	0

FIGURE 8.8: Total mappings and correct mappings from $\sigma_n = [0.5..1]$ using DbMN_6, where red depicts a lower number of mappings found, and green a higher number of mappings found.

Figure 8.8 also shows that the alignment at 0.625 for these selected ontologies consists of an alignment with three mappings. Following from 0.625, the ontologies O= cmt ➟ O'= confof and O= conference ➟ O'= confof are both unable to find alignments from ϵ_n = 0.650 onwards using DbMN_5 and both at ϵ_n = 0.650 find no alignment. The remaining ontologies no longer find an alignment from ϵ_n = [0.725..0.750]. This illustrates that a neighbourhood threshold of ϵ_n = 0.750 is too high for any mappings to be mutually accepted by the agents, resulting in no alignments generated by DbNM_6.

These results suggest that the same alignments are generated in both variants from DbNM_5 and DbNM_6. This similarity can be seen in an example taken from the ontologies O= cmt ➟ O'= conference show in Figure 8.9 from the values of ϵ_n= [0..1].

	n_threshold 0.000, w_0.25	n_threshold 0.025, w_0.25	n_threshold 0.500, w_0.25	n_threshold 0.550, w_0.25	n_threshold 0.600, w_0.25	n_threshold 0.625, w_0.25	n_threshold 0.650, w_0.25	n_threshold 0.675, w_0.25	n_threshold 0.700, w_0.25	n_threshold 0.725, w_0.25
SYSTEM V5 TOTAL MAPS found:	29	26	24	16	15	8	4	2	1	0
SYSTEM V6 TOTAL MAPS found:	29	26	24	16	15	8	4	2	1	0
SYSTEM V5 TOTAL MAPS found INC IN PS:	6	5	4	4	4	3	2	1	1	0
SYSTEM V6 TOTAL MAPS found INC IN PS:	6	5	4	4	4	3	2	1	1	0

FIGURE 8.9: Heat map of total mappings found, illustrating only the decreased values for the sake of visual clarity, O= cmt➟ O'= conference

Similar to DbMN_5 these results show that DbMN_6 can find a meaningful alignment between two ontologies and maintain partial sharing of the knowledge base O'. This is limited to a bound on the neighbourhood threshold value of [0..0.625] for all the dataset pairs represented. This illustrates that at ϵ_n= 0.650 DbMN_6 is overfitting for both the precision and recall values, illustrating that no *correct* mappings are included in the alignments resulting in the curves reducing to 0 at this threshold point.

[Q.2] *Can the variant DbMN_6 generate a plausible alignment and maintain a level of privacy?*

Over the ontologies presented in this evaluation a meaningful alignment is found between a ϵ_n bound of [0..0.750]. From ϵ_n= [0.750..1] (as with DbMN_5) the DbMN_6 variant finds no meaningful alignment across any of the dataset pairs.

From these results *Q.2* addresses the number of *private* concepts from O' unshared throughout the iteration of the dialogue over the proponent's signature. The variable setting, *V.3* discusses that DbMN_6 includes all the concepts from ontology O in the signature to be mapped resulting in no privacy for the proponent. However, the opponent is able to maintain a level of privacy throughout the dialogue process for DbMN_6. This privacy of O' can be seen in the results presented in Table 8.7 where the unshared concepts of the ontology are presented.

A level of privacy can be maintained by the opponent agent within this variant of the DbMN across the full bound of the neighbourhood threshold. When this threshold is set to its minimum bound of $\epsilon_n = 0$, the opponent agent using the DbMN is successful in generating an alignment, and sharing a maximum of 82.05% of the ontology. As hypothesised, when the neighbourhood threshold increase, there is a decrease in the sharing of the opponents ontology of 5.43% using the ontologies O = conference ➟ O' confof.

TABLE 8.7: Table presenting the level of privacy for the opponent agent's ontology O' for DbMN_6

O -$O\prime$	$e \in O'$ shared at $\epsilon_n = 0$	$e \in O'$ shared at $\epsilon_n = 1$	total no. entities $e \in O'$	$e \in O'$ unshared between ϵ_n0-1	% of $O\prime$ shared at $\epsilon_n = 0$	% of $O\prime$ shared at $\epsilon_n = 1$
cmt-conference	26	25	60	1	43.33%	41.67%
cmt-confof	23	22	39	1	58.97%	56.41%
cmt-edas	34	31	104	3	32.69%	29.81%
cmt-ekaw	32	27	74	5	43.24%	36.49%
cmt-sigkdd	24	23	50	1	48%	46%
conference-confof	32	30	39	2	82.05%	76.92%
conference-edas	50	42	104	8	48.08%	40.38%
conference-ekaw	50	45	74	5	67.57%	60.81%
conference-sigkdd	33	30	50	3	66%	60%
confof-edas	39	36	104	3	37.5%	34.62%
confof-ekaw	35	32	74	3	47.3%	43.24%
confof-sigkdd	26	26	50	0	52%	52%
edas-ekaw	51	46	74	5	68.92%	62.16%
edas-sigkdd	39	35	50	4	78%	70%
ekaw-sigkdd	37	33	50	4	74%	66%

It can be seen in Table 8.7 that from the values of ϵ_n=[0..1] over the ontologies used in this evaluation, there is a reduction in the number of concepts shared from ontology O' ranging from 0 concepts unshared to 8 concepts unshared.

In comparison to DbMN_5, there are no examples between ϵ_n=[0..1] where more concepts are shared, however an individual instance where there is an equal number shared between ϵ_n=[0..1]:

O=confof ➟ O'=sigkdd

This correction from the previous variant is explained by the difference in the sharing strategy used in DbMN_6. Here the full neighbourhood of a concept is shared in one *testify* move by the sender rather than individually in DbMN_5.

This results in full neighbourhood of a target concept being shared where the subject of the triple is equivalent to the concept under negotiation. The premise Pr (which is required to exceed the threshold value) thus comprises multiple triples shared in one iteration of the *testify* move. This sharing strategy results in more concepts being shared at ϵ_n=0 in comparison to DbMN_5 illustrated in Table 8.8.

TABLE 8.8: Comparing % of concepts shared at ϵ_n= 0 between DbMN_5 and DbMN_6.

	O = cmt					conference				confof			edas		ekaw
	O'= conference	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= edas	O'= ekaw	O'= sigkdd	O'= ekaw	O'= sigkdd	O'= sigkdd
DbMN V_6	43.3	59	**32.7**	**43.2**	**48**	**82.1**	48.1	67.6	**66**	**37.5**	47.3	**52**	68.9	78	74
DbMN V_5	43.3	59	32	40.5	42	79.5	48.1	67.6	64	35.6	48.7	48.7	68.9	78	74

Overall, when using DbMN_6 it can be seen that there is an increase in the number of concepts shared at the value of ϵ_n= 0 (illustrated in Table 8.8) where in the following cases, there are more concepts shared by DbMN_6 at ϵ_n= 0 than DbMN_5:

O = cmt ➟ O' = edas

O = cmt ➟ O' = ekaw

O = cmt ➟ O' = sigkdd

O = conference ➟ O' = sigkdd

O = conference ➟ O' = confof

O = confof ➟ O' = sigkdd

O = confof ➟ O' = edas

This increase in sharing at ϵ_n= 0 can be explained with the varied agent strategy, meaning that at the first *testify* move the opponent agent will share the full neighbourhood of the concept proposed at the *propose* move, rather than individually as done in DbMN_5. Table 8.7 also illustrates the percentage of O' which is shared when using DbMN_5. Across all the ontologies, there is maximum of 82.05% of concepts in ontology O' and a minimum of 32.69% shared at ϵ_n= 0, and a maximum of 76.92% of concepts in ontology O' and a minimum of 29.81% shared at ϵ_n= 1.

This sharing reflects the ability of this variant of the approach to allowing the opponent to maintain a level of privacy by not sharing their full ontology O' over the dialogue for the proponents signature to be mapped, taking into consideration the adjusted sharing mechanism in this version of the dialogue.

[Q.3] *What are the influences of the threshold levels on the accuracy and correctness of the alignment*

The neighbourhood threshold (defined in Section 8.2), has been evaluated in isolation to investigate the importance of neighbourhood in the generation of a mapping. When this neighbourhood threshold is increased from [0..1] across all of the results for all ontology pairs, there is a negative correlation found over the number of mappings accepted by the agents into the alignment (similar to DbMN_5). The DbMN_6 version of the approach finds a bound of this neighbourhood threshold between [0..0.750]. This is the point where no further mappings are included in the alignment. This is a result of the approach to prune out the potentially *incorrect* mappings as candidate mappings, where the neighbourhood similarity given the supporting premise does not exceed the threshold levels. This; however, illustrates that the approach is over fitting the alignments and pruning out *correct* mappings. This gives DbMN_6 an upper bound of 0.750 where this is providing an outputted alignment.

The results of the influence of this varying neighbourhood threshold were investigated using the same evaluation methods as DbMN_5, with the adjustment of the neighbourhood sharing to multiple triples in a single *testify* move. With these restrictions the outputted precision and recall values for the DbMN_6 dataset pairs produced the same figures as DbMN_5. These results for DbMN_6 have been summarised in Table 8.9 for all the dataset pairs used, illustrating the highest and lowest bounds of these values, and the overall total mappings and number of mappings in the benchmark standard for the pairs of ontologies.

[Q.4] *What are the influences of the neighbourhood sharing order on the alignment?*

The final experimental question explored the choice of ranking function used by the agents determining which triple is shared in a *testify* move throughout the dialogue exchange. In contrast to DbMN_5, DbMN_6 does not rank these triples or use an ordering, as they are all shared in a single *testify* move.

In the evaluation of DbMN_6, it was found that the overall alignments generated were identical to those found in DbMN_5 over the same range of ϵ_n threshold values. Although this was expected however, the results present a difference in the sharing of the triples. This similarity of the alignments generated by the two variants can be seen in the heat map illustrated in Figure 8.9 and discussed in *Q.2*.

This alternate sharing is illustrated in Figure 8.10 using the ontologies O=cmt ➟ O'=edas, with the value of ϵ_n =0.700. The concept initiated by the proponent is *Paper*; this figure shows the opponent agent sharing a single triple for their concept *Paper* in DbMN_5, and both triples i.e. the full neighbourhood in the same *testify* move using DbMN_6. It illustrates that using both variants the same information is shared, however with the DbMN_6 it is shared in a single *testify* move.

	Dialogue Phase	**Move**
DbMN_5	OPEN	<Proponent initiate cmt#Paper null null>
	PROPOSE	<Oponent propose cmt#Paper edas#Paper null>
	CONFIRM	<Proponent justify cmt#Paper edas#Paper null>
	CONFIRM	<Oponent testify cmt#Paper edas#Paper <edas#Paper edas#isWrittenBy edas#Author>>
	CONFIRM	<Proponent justify cmt#Paper edas#Paper null>
	CONFIRM	<Oponent testify cmt#Paper edas#Paper <edas#Paper IS_A edas#Document>>
	CONFIRM	<Proponent assert cmt#Paper edas#Paper [(<edas#Paper IS_A edas#Document> <cmt#Paper IS_A cmt#Document>)]>
	CONFIRM	<Oponent accept cmt#Paper edas#Paper [(<cmt#Paper IS_A cmt#Document> <edas#Paper IS_A edas#Document>)]>
	CLOSE	Dialogue Ended On Initated Concept
DbMN_6	OPEN	<Proponent initiate cmt#Paper null null>
	PROPOSE	<Oponent propose cmt#Paper edas#Paper null>
	CONFIRM	<Proponent justify cmt#Paper edas#Paper null>
	CONFIRM	<Oponent testify cmt#Paper edas#Paper [<edas#Paper edas#isWrittenBy edas#Author> <edas#Paper IS_A edas#Document>]>
	CONFIRM	<Proponent assert cmt#Paper edas#Paper [(<edas#Paper IS_A edas#Document> <cmt#Paper IS_A cmt#Document>)]>
	CONFIRM	<Oponent accept cmt#Paper edas#Paper [(<cmt#Paper IS_A cmt#Document> <edas#Paper IS_A edas#Document>)]>
	CLOSE	Dialogue Ended On Initated Concept

FIGURE 8.10: Figure illustration the sharing of a full neighbourhood in DbMN_6, in comparison to the single sharing in DbMN_5.

TABLE 8.9: DbMN_6 Results summary table

	O = cmt					conference				confof			edas		ekaw
	O'= conference	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= edas	O'= ekaw	O'= sigkdd	O'= ekaw	O'= sigkdd	O'= sigkdd
Highest Precision	ϵ_n=.700 1	ϵ_n=.650 0.67	ϵ_n=.650 0.63	ϵ_n=.700 0.40	ϵ_n=.450 0.31	ϵ_n=.650 0.67	ϵ_n=.725 0.5	ϵ_n=.650 0.33	ϵ_n=.650 0.67	ϵ_n=.700 1	ϵ_n=.675 1	ϵ_n=.625 0.17	ϵ_n=.725 0.25	ϵ_n=.750 0.2	ϵ_n=.675 0.6
Highest Recall	ϵ_n=.000 0.67	ϵ_n=.000 1	ϵ_n=.000 1	ϵ_n=.000 0.71	ϵ_n=.000 0.89	ϵ_n=.000 0.80	ϵ_n=.000 0.89	ϵ_n=.000 0.87	ϵ_n=.000 0.82	ϵ_n=.000 1	ϵ_n=.000 0.75	ϵ_n=.000 1	ϵ_n=.000 0.88	ϵ_n=.000 1	ϵ_n=.000 1
Lowest Precision	ϵ_n=.725 0	ϵ_n=.675 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.675 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.650 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0
Lowest Recall	ϵ_n=.725 0	ϵ_n=.675 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.675 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.650 0	ϵ_n=.750 0	ϵ_n=.750 0	ϵ_n=.750 0
Most Mappings	ϵ_n=.000 29	ϵ_n=.000 29	ϵ_n=.000 29	ϵ_n=.000 29	ϵ_n=.000 29	ϵ_n=.000 59	ϵ_n=.000 59	ϵ_n=.000 59	ϵ_n=.000 59	ϵ_n=.000 38	ϵ_n=.000 38	ϵ_n=.000 38	ϵ_n=.000 103	ϵ_n=.000 103	ϵ_n=.000 73
Most Correct Mappings	ϵ_n=.000 6	ϵ_n=.000 5	ϵ_n=.000 8	ϵ_n=.000 5	ϵ_n=.000 8	ϵ_n=.000 8	ϵ_n=.000 8	ϵ_n=.000 13	ϵ_n=.000 9	ϵ_n=.000 10	ϵ_n=.000 12	ϵ_n=.000 4	ϵ_n=.000 14	ϵ_n=.000 7	ϵ_n=.000 7
Total Mappings	63	50	58	59	50	93	117	110	98	75	72	67	167	136	122
Benchmark Mappings	9	5	8	7	9	10	9	15	11	10	16	4	16	7	7

[Summary]

To summarise the DbMN_6 evaluation, the variation in sharing strategy generates the same alignments as DbMN_5. However, the difference in this variant is seen in the sharing iterations, at the lowest neighbourhood threshold of 0.

Overall DbMN_6 finds meaningful alignments between the same bounds as DbMN_5, thus supports the hypothesis posed. This variant, in comparison with DbMN_5, shares less of the ontology throughout the neighbourhood threshold increments, meaning that the sharing is higher at 0, yet still finds the same alignments as DbMN_5. This batch sharing approach in DbMN_6 still shared the same concepts as the strategy in DbMN_5 and provided the same alignment results. Furthermore, the dialogue process only completes one iteration of the *testify-justify* loop as predicted. This variant demonstrates that there is no further privacy gained by sharing the concept locality individually. However, the number of moves used in dialogue process is reduced.

Overall DbMN_6 supports the hypothesis that a meaningful alignment can be found using the partial sharing in the a dialogical protocol across a set of incremented neighbourhood threshold values. The variables detailed above were selected in order to evaluate the influences of these variables over the alignments generated by the approach across the incremented neighbourhood threshold levels, to find the optimum range over its precision and recall when compared to the platinum standard reference alignment. This would then provide an optimum range where the threshold bound found the most accurate alignments created.

The rationale behind this decision was that as the neighbourhood threshold increased, this variant would find it harder to accept weaker and potentially *incorrect* mappings and the final alignment would reduce in the number of mappings found, from the threshold value set from [0..1], resulting in a more accurate alignment towards the value of 1.

Furthermore it was hypothesised that by restricting mappings to one to one mappings this would reduce the amount of the ontology O' shared throughout the dialogue. Using the variables detailed above, the final variant DbMN_7, was evaluated to examine the experimental questions *Q.1..Q.5* detailed in Section 8.2.

8.3.3 Empirical evaluation of approach; DbMN_7

This section describes the experiments used to to evaluate the performance of DbMN_7 in generating an alignment over the range of neighbourhood threshold values. The aim of this empirical evaluation it to compare its performance with that of the previous variants, to answer the following four experimental questions, [Q.1..Q.4], described below.

[Q.1] *Can a plausible meaningful alignment be found, that compares to a benchmark standard, when two agents engage in a dialogical approach to ontology matching, rather than sharing their knowledge bases in full, a priori?*

From the ontologies used in this implementation, a meaningful alignment can be found across a range of neighbourhood thresholds ϵ_n= [0..0.750]. The results suggest that there is a similar decrease in the number of mappings accepted to an alignment at ϵ_n= 0.750, suggesting that the approach at this point is over fitting the alignments and failing to accept candidate correspondences. Figure 8.11 plots the number of total

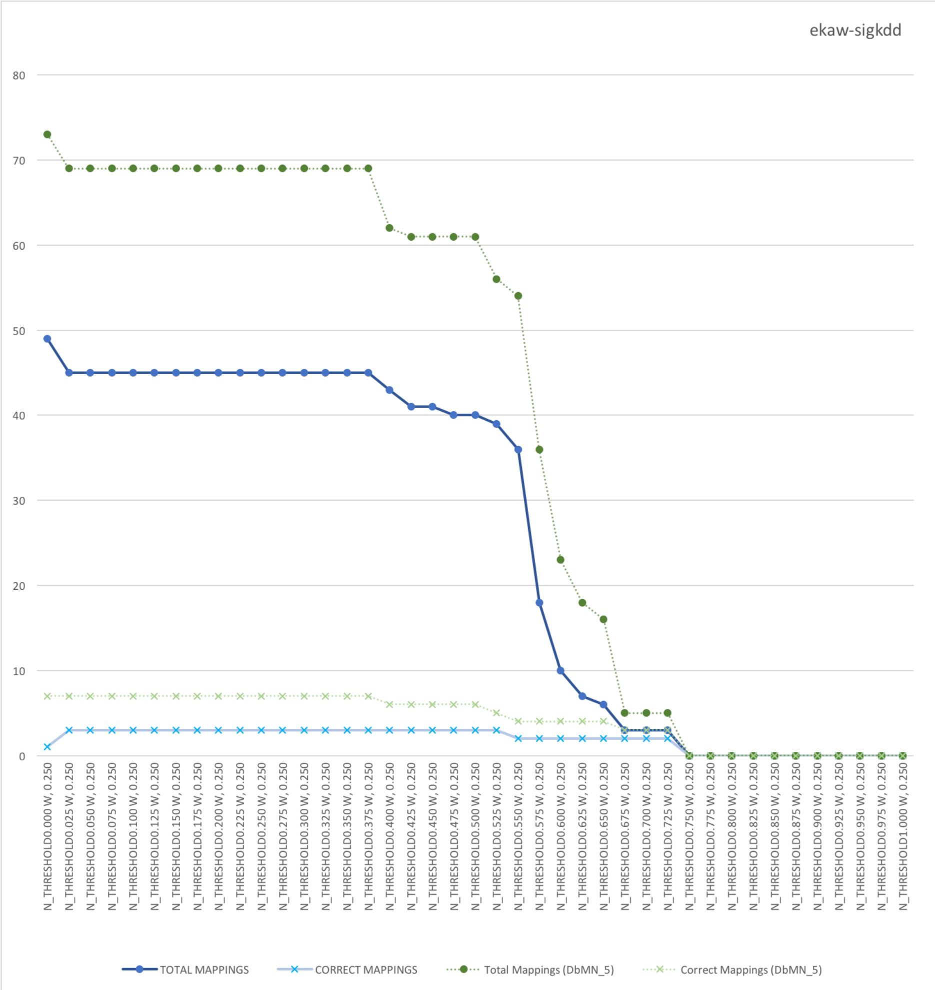


FIGURE 8.11: O = ekaw, O' = sigkdd total and correct Mappings

correct mappings and mappings found by DbMN_7, in comparison to DbMN_5 using the following ontologies:

O = ekaw ➟ O' = sigkdd.

Overall, with a restriction of 1:1 mappings, DbMN_7 finds fewer total and *correct* mappings than DbMN_5 across the range of neighbourhood thresholds tested ϵ_n (the full results of the individual ontologies can be found in Appendix A). Using the ontologies O = ekaw ➟ O' = sigkdd at ϵ_n= 0, DbMN_7 finds 49 total mappings and 3 *correct* mappings at ϵ_n= 0.025.

Due to the 1:1 mapping restriction, the DbMN_7 finds fewer *incorrect* mappings than DbMN_5, which is a positive finding for the approach, thus generating more accurate alignments. However, this variant also finds fewer *correct* mappings as a result of this restriction. Whist the variant finds a meaningful *correct* alignment between the two ontologies, this is lower than the previous two.

[Q.2] *Can the approach generate a plausible alignment, and maintain a level of privacy?*

In this dialogue variant, the type of alignment generated is restricted to only include injective (i.e 1:1) mappings, such that any concept already mapped can not be included in a *propose* move, allowing for a decrease in the level of disclosure of the opponent ontology O'. This illustrated in Table 8.10, which lists the level of disclosure of concepts when evaluating the ontology pair:

O = edas ➟ O' = ekaw

TABLE 8.10: Table presenting the level of privacy for the opponent agent's ontology O' in DbMN_7

O -$O_\prime$	$e \in O'$ shared at ϵ_n= 0	$e \in O'$ shared at ϵ_n= 1	total no. entities $e \in O'$	$e \in O'$ unshared between ϵ_n0-1	% of $O_\prime$ shared at ϵ_n= 0	% of $O_\prime$ shared at ϵ_n= 1
cmt-conference	40	25	60	15	67%	41.67%
cmt-confof	31	22	39	9	79.49%	56.41%
cmt-edas	40	31	104	9	38.46%	29.81%
cmt-ekaw	42	27	74	15	56.76%	36.49%
cmt-sigkdd	33	23	50	10	66%	46%
conference-confof	38	30	39	8	97.44%	76.92%
conference-edas	75	42	104	33	72.12%	40.38%
conference-ekaw	66	45	74	21	89.19%	60.81%
conference-sigkdd	49	30	50	19	98%	60%
confof-edas	49	37	104	12	47%	36%
confof-ekaw	47	33	74	14	64%	45%
confof-sigkdd	41	27	50	14	82%	54%
edas-ekaw	73	46	74	27	98.65%	62.16%
edas-sigkdd	49	35	50	14	98%	70%
ekaw-sigkdd	49	33	50	16	98%	66%

A decrease in the number of concepts shared was observed from ontology O' from 73 concepts to 46 concepts, where the neighbourhood threshold is set between ϵ_n= [0..1]. This shows an improved decrease in disclosure, when compared to the previous two variants.

This table shows that with a higher neighbourhood threshold, fewer of the entities were shared. This could be explained by the opponent disclosing more concepts at ϵ_n = 0 (in DbMN_7) than at the same threshold in DbMN_5 and 6, suggesting that there is more room for reducing this initial disclosure. In Table 8.11, the comparison of the percentage of disclosure of the opponent's ontology is presented between neighbourhood threshold values ϵ_n = [0..1].

This illustrates that DbMN_7 (at ϵ_n= 0) shares a minimum of 5.8% more concepts from O' (using the ontologies O = cmt ➟ O' = edas) than the previous two variants. This could be a result of the 1:1 restriction, where the opponent agent is restricted to proposing concepts that have previously not already been mapped. Overall, in answering the research question posed by *Q3*, DbMN_7 is able to find an alignment between the two agents whilst preserving undisclosed concepts from the opponent's ontology.

TABLE 8.11: Comparing the % of concepts shared at ϵ_n= [0..1] between DbMN_5, DbMN_6 and DbMN_7.

	O = cmt					conference				confof			edas		ekaw
	O'= conference	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= confof	O'= edas	O'= ekaw	O'= sigkdd	O'= edas	O'= ekaw	O'= sigkdd	O'= ekaw	O'= sigkdd	O'= sigkdd
ϵ_n=0															
DbMN V_7	67	79.5	38.5	56.7	66	97.4	72.1	89.2	98	47	64	82	98.7	98	98
DbMN V_6	43.3	59	32.7	43.2	48	82.1	48.1	67.6	66	37.5	47.3	52	68.9	78	74
DbMN V_5	43.3	59	32	40.5	42	79.5	48.1	67.6	64	35.6	48.7	48.7	68.9	78	74
ϵ_n=1															
DbMN V_7	41.7	56.4	29.8	36.5	46	76.9	40.4	60.8	60	36	45	54	62.2	70	66
DbMN V_6	41.7	56.4	29.8	36.5	46	76.9	40.4	60.8	60	36	45	54	62.2	70	66
DbMN V_5	41.7	56.4	29.8	36.5	46	76.9	40.4	60.8	60	36	45	54	62.2	70	66

[Q.3] *What are the influences of the threshold levels on the accuracy and correctness of the alignment?*

DbMN_7 finds a general positive correlation between the incrementation of the neighbourhood threshold (ϵ_n) over the range [0..1], and the precision values and a negative correlation between the ϵ_n and the recall values.

- **Influence of the neighbourhood threshold value (ϵ_n) on the precision of the outputted alignment.**

  Figure 8.12 illustrates the positive correlation between the neighbourhood similarity and the precision of the alignments generated. The precision curves for all the ontologies demonstrate a positive correlation of the precision from ϵ_n= [0..1]. The neighbourhood threshold ϵ_n= 0.725, presents the upper bound where an alignment is still successfully generated before the approach fully overfits the alignments. This figure shows that no mappings for any of the ontologies are found in the alignments from ϵ_n = [0.750..1], and have been removed for clarity. It can be seen in Figure 8.12 that the highest precision value for the alignments generated by the iterated neighbourhood threshold values is 1, over the following ontologies:

  O = ekaw ➟ O' = sigkdd O = conference ➟ O' = ekaw
  O = cmt ➟ O' = confof O = cmt ➟ O' = conference
  O = confof ➟ O' = ekaw, O = confof ➟ O' = edas

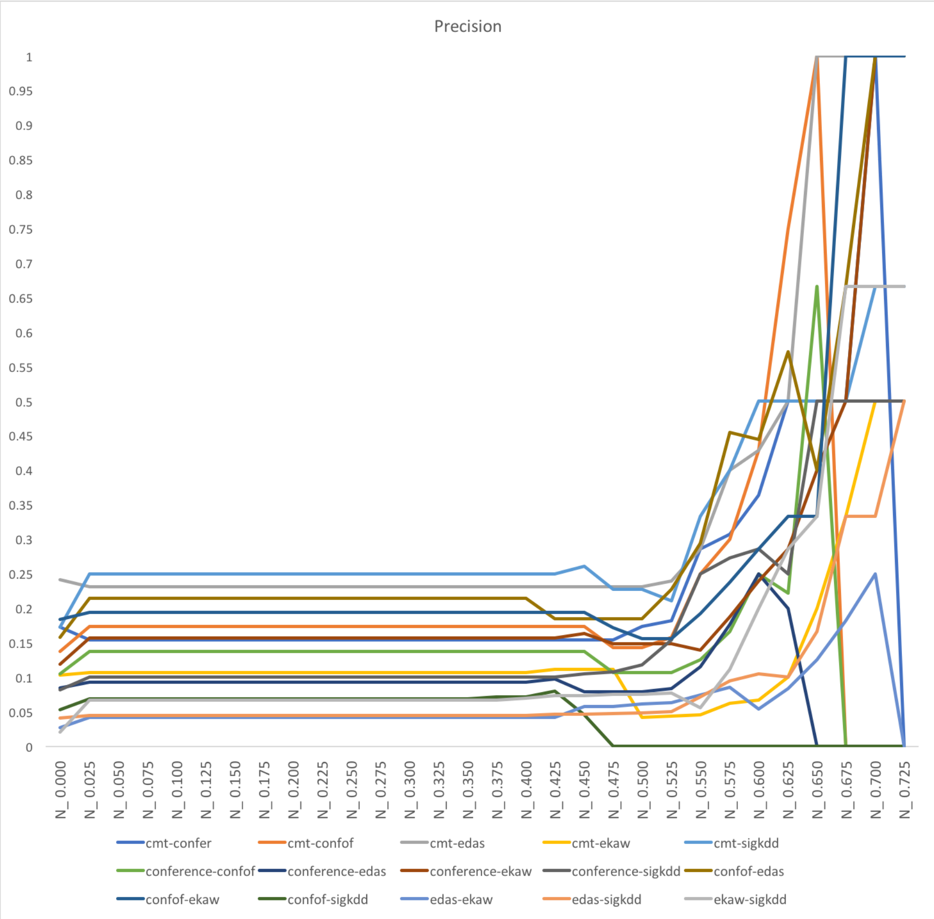


FIGURE 8.12: Precision curves for all the ontologies using DbMN_7

With these results, DbMN_7 finds a higher precision over more ontology pairs than the two previous variants, suggesting that the 1:1 mapping restriction finds a more accurate alignment between ϵ_n =[0.625..0.700].

The Figure 8.12 also shows the precision curves for the ontologies O = cmt ➟ O' = sigkdd, which decreases between ϵ_n =[0.475..0.525] before an increase at 0.550. This is due to the fact that when $\epsilon_n = 0.550$, 7 *incorrect* mappings are prune; whilst maintaining the number of *correct* mappings found is maintained, thus creating a higher level of precision.

The decrease in the precision between ϵ_n = [0.475..0.525] is a result of the approach failing to find a single mapping appearing in the platinum standard. Between ϵ_n =[0.450..0.475] the mapping $\langle Document, Document, \equiv\rangle$ is not found and between ϵ_n =[0.500..0.525] the mapping $\langle Person, Person, \equiv\rangle$ is not found.

These mappings are not included in the alignment as the agents fail to obtain the necessary support in the premise of the assertions needed to accept these mappings. The increase in precision is a result of the variant removing *incorrect* mappings; this suggests that the approach working as expected.

The Figure 8.12 also presents the precision curves for the ontologies O = confof ➟ O' = sigkdd, which is the lowest between of the ontologies. This curve illustrates that DbMN_7 fails to find mappings in an alignment when $\epsilon_n = 0.475$, which is the earliest point of overfitting. This could be due to the premise no longer meeting the neighbourhood bound, and therefore no mappings can be accepted.

- **Influence of the neighbourhood threshold value (ϵ_n) on the recall of the outputted alignment.**

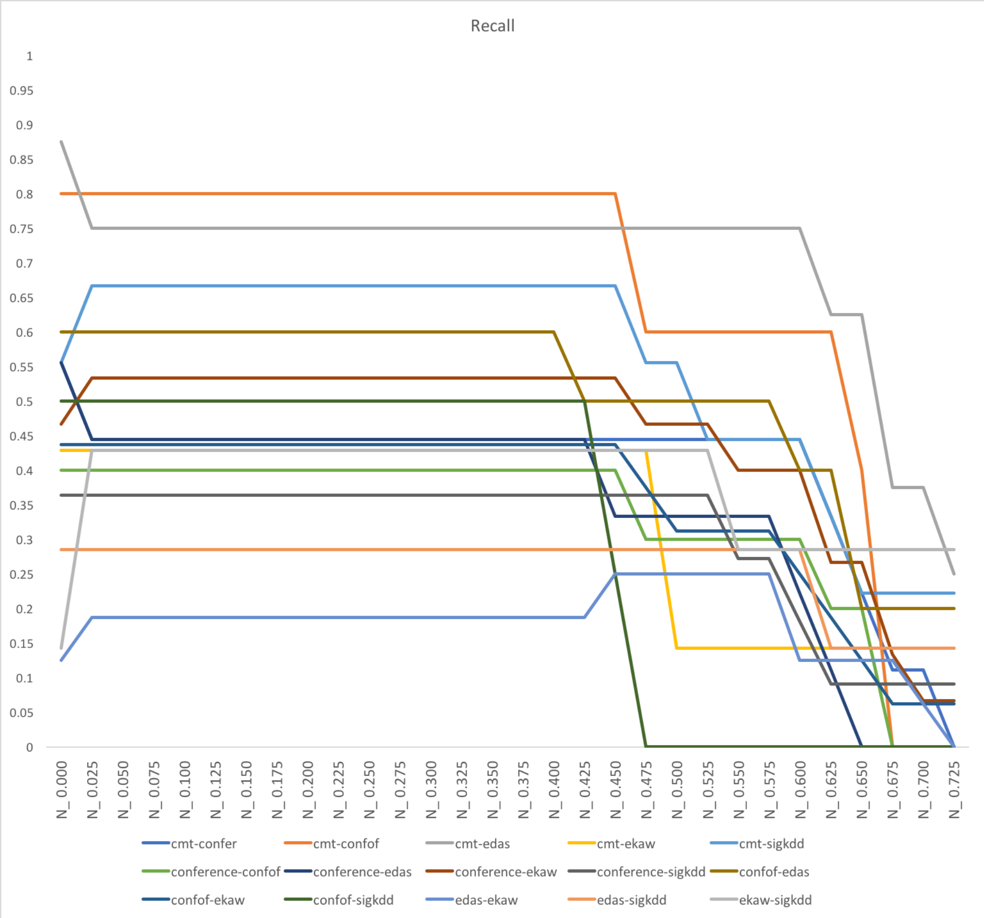
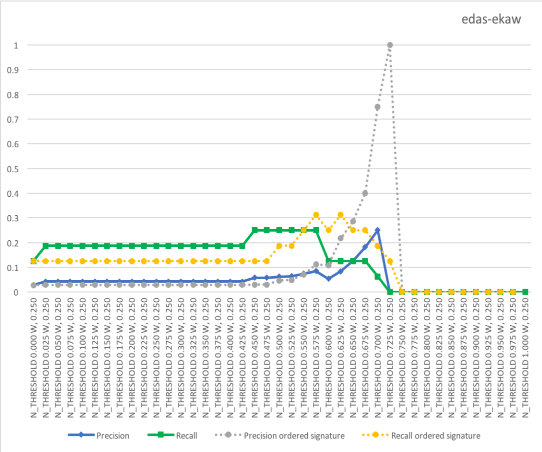


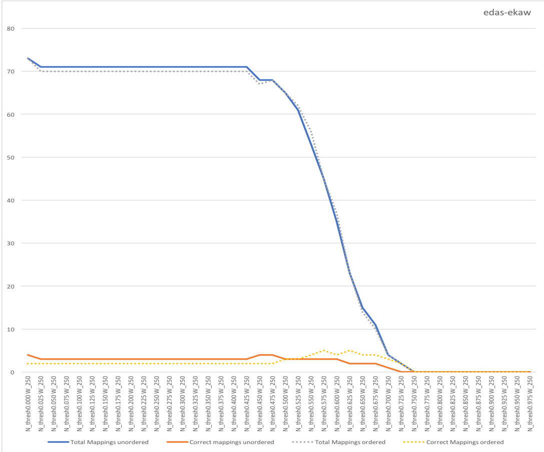
FIGURE 8.13: Recall curves for all the ontologies using DbMN_7

DbMN_7 also finds a strong negative correlation across all the evaluated ontologies for the range, ϵ_n=[0..1] over the recall value. These values can be seen in the following results in Figure 8.13.

Figure 8.13 illustrates the highest recall value is 0.9 presenting an overall decrease in the recall values found by this variant to those found by DbMN_5 and DbMN_6 with a highest recall value of 1. This recall value decreases gradually as the neighbourhood threshold increases in the range of ϵ_n= [0..0.750], showing that the recall can no longer find a meaningful alignment between the ontologies. This overall



(a) O = edas, O' = ekaw precision and recall



(b) O = edas, O' = ekaw total and correct mappings

FIGURE 8.14: For the ontologies O = edas, O' = ekaw, (a) presents the precision and recall and (b) presents the total mapping.

lower bound of recall values for DbMN_7 could be explained with the 1:1 restriction allowing it to find fewer *incorrect* mappings in the alignments between the ontologies.

[Q.4] *What are the influences of the neighbourhood sharing order on the alignment?*

This experimental question uses the weights of the ranking function used by the agents to determine how the agents share a chosen triple in a *testify* move within the dialogue exchange. In contrast to DbMN_6, DbMN_7 works the same way as DbMN_5 in sharing the neighbourhood of a concept individually in an iteration of *testify* moves. This varied the individual weight of the ranking function (*Subsumption*, *Rarity*, *Connectivity* and *Popularity*), and compared the resulting ranking of concepts with that when the order was alphabetic. These two ranking orders were evaluated over a range of neighbourhood thresholds. The initial set of experiments evaluated using DbMN_7 showed no difference on the outputted alignments over each increment of the neighbourhood similarity thresholds, using either the ranking method weighted evenly, or using the 0 or 1 weightings, to using an alphabetic ranking. This suggested that the ranking had no influence on the outputted alignments, and for the purposes of the evaluation of the results for DbMN_7, only the equally weighted ranking values have been used for the empirical evaluation of this approach.

[Q.5] *What are the influences of the signature order on the accuracy and correctness of the alignment?*

The final experimental question investigates the influence of the ordering of the concepts in the signature Σ^t on the alignments generated. The previous variants utilise a random ordering for the concepts in the signature: however due to the 1:1 restriction, this ordering was investigated in this variant, as the order will distinguish which concepts can be proposed. If the concept has already been mapped it cannot be reused at a *propose*

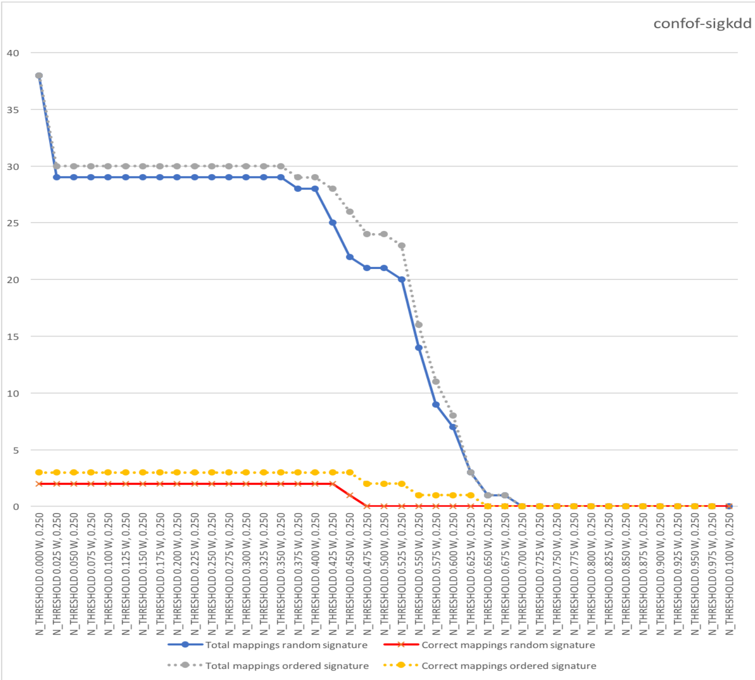


FIGURE 8.15: O = confof, O' = sigkdd Precision and Recall

move, therefore the signature order determines what concepts the opponent agent can propose. The influence of the signature ordered was investigated, using both a randomly generated order and an alphabetic order to examine the influences of these differences on the alignment generated with a 1:1 mapping restriction.

Figure 8.14 (a) illustrates difference between the precision and recall in terms of the alignments produced by the approach, using an alphabetically ordered or randomly generated signature containing the concepts from the proponent's ontology O. It can be seen from the graph that using the ordered signature, results in an increase in precision for the alignments when using the thresholds between ϵ_n=[0.550..0.750]. This also shows an improvement on the performance of the randomly generated signature increasing the precision value to 1. The recall value fluctuates when the neighbourhood threshold is increase in the range ϵ_n=[0.475..0.750] which could be explained with the positive correlation of *correct* mappings over the same neighbourhood increments as those found in this increased recall value illustrated in Figure 8.14 (b).

Figure 8.15 illustrates a more consistent set of results for both the precision and recall and the total and correct number of mappings found over the ordered and unordered signature. In this figure the precision and recall values found over the ordered signature for the same ontologies is higher across all the neighbourhood threshold bounds from ϵ_n=[0..1].

Figure 8.16 illustrates that the total and number of *correct* mappings found is also higher using this ordered signature. This consistent improvement in the results for DbMN_7 (using the ordered over a randomly generated signature) shows that there is

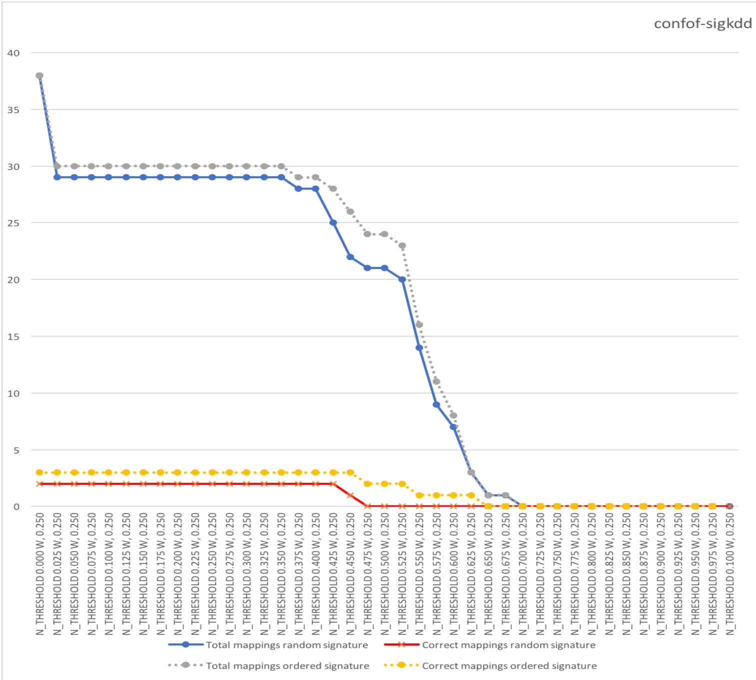


FIGURE 8.16: O = confof, O' = sigkdd total and *correct* Mappings

inconsistencies in the performance across the ontologies and could be a result of the ontology pairs used in this evaluation, but it is inherent that the signature order has an influence on the alignments generated and accepted.

The influence of the signature order can be seen in Figure 8.17 using the ontologies, O = cmt ➟ O' = conference for ϵ_n= 0 and the opponent concept $Presentation \in O'$. Using the alphabetically ordered signature, the generated alignment includes the mapping $\langle PaperFullVersion, Presentation \equiv\rangle$. Using the randomly ordered signature, the alignment generated includes the mapping $\langle Preference, Presentation \equiv\rangle$. This difference in the mappings for $Presentation \in O'$ is a result of the ordering of the proponent agent's signature.

[Summary]

To summaries DbMN_7, with its variation in sharing strategy, and restricting the type of mappings to 1:1, it is able to find meaningful alignments between ontologies when the neighbourhood threshold is varied in the range of [0..0.750], and thus supports the hypothesis posed. In comparison with DbMN_6 and DbMN_5, this variation using the 1:1 mappings shares more concepts within the ontology when the neighbourhood threshold is 0, converging on the same sharing quantities as the previous variants when the same threshold is 1. DbMN_7 due to its 1:1 mapping restrictions also prunes out more *incorrect* mappings at the lowest neighbourhood threshold value, and finds overall more accurate alignments.

	sig concept number	Move
DbMN_7 ordered signature	19	<Alice initiate cmt#PaperFullVersion null null>
	19	<Bob propose cmt#PaperFullVersion conference#Presentation null>
	19	<Alice justify cmt#PaperFullVersion conference#Presentation null>
	19	<Bob testify cmt#PaperFullVersion conference#Presentation <conference#Presentation conference#is_given_by conference#Active_conference_participant>>
	19	<Alice assert cmt#PaperFullVersion conference#Presentation [(<conference#Presentation conference#is_given_by conference#Active_conference_participant>
	19	<Bob accept cmt#PaperFullVersion conference#Presentation [(<cmt#ProgramCommitteeMember IS_A cmt#ConferenceMember> <conference#Program_committee IS_A conference#Committee>)]>
	19	Dialogue Ended On Initated Concept
	21	<Alice initiate cmt#Preference null null>
	21	<Bob propose cmt#Preference conference#Review_preference null>
	21	<Alice justify cmt#Preference conference#Review_preference null>
	21	<Bob testify cmt#Preference conference#Review_preference null>
	21	<Alice assert cmt#Preference conference#Review_preference [(<conference#Poster IS_A conference#Conference_contribution> <cmt#ProgramCommitteeMember IS_A
	21	<Bob accept cmt#Preference conference#Review_preference [(<cmt#ProgramCommitteeMember IS_A cmt#ConferenceMember> <conference#Program_committee IS_A conference#Committee>)]>
	21	Dialogue Ended On Initated Concept
DbMN_7 random signature	7	<Alice initiate cmt#Preference null null>
	7	<Bob propose cmt#Preference conference#Presentation null>
	7	<Alice justify cmt#Preference conference#Presentation null>
	7	<Bob testify cmt#Preference conference#Presentation <conference#Presentation conference#is_given_by conference#Active_conference_participant>>
	7	<Alice assert cmt#Preference conference#Presentation [(<conference#Presentation conference#is_given_by conference#Active_conference_participant>
	7	<Bob accept cmt#Preference conference#Presentation [(<cmt#ProgramCommitteeMember IS_A cmt#ConferenceMember> <conference#Program_committee IS_A conference#Committee>)]>
	7	Dialogue Ended On Initated Concept
	14	<Alice initiate cmt#PaperFullVersion null null>
	14	<Bob propose cmt#PaperFullVersion conference#Paid_applicant null>
	14	<Alice justify cmt#PaperFullVersion conference#Paid_applicant null>
	14	<Bob testify cmt#PaperFullVersion conference#Paid_applicant <conference#Paid_applicant IS_A conference#Registeered_applicant>>
	14	<Alice assert cmt#PaperFullVersion conference#Paid_applicant [(<conference#Paid_applicant IS_A conference#Registeered_applicant> <cmt#Meta-Reviewer IS_A cmt#Reviewer>)]>
	14	<Bob accept cmt#PaperFullVersion conference#Paid_applicant [(<cmt#Meta-Reviewer IS_A cmt#Reviewer> <conference#Rejected_contribution IS_A conference#Reviewed_contribution>)]>
	14	Dialogue Ended On Initated Concept

FIGURE 8.17: Figure illustration the signature order effecting the alignments.

Furthermore, the ordered signature of the proponent agent, has an effect on the alignments generated as a result of the 1:1 mappings. In some cases (seen in the Appendix A) this has proved to produce better precision and recall results; however in other cases, this order restriction on the signature has had a negative effect on the precision and recall of the alignment generated, producing lower results. This could be due to the more accurate concepts mapped leaving weaker or lower similarity concepts available for candidate mappings, therefore producing a lower confidence mappings in the alignment. It can be seen that there remains the bound of the neighbourhood threshold level, which at ϵ_n= 0.750 fails to produce an alignment between the ontologies.

Overall the results for DbMN_7 supports the hypothesis that a meaningful alignment can be found using the partial sharing in the approach, across a set of incremented neighbourhood threshold values, and both over an ordered signature and a random signature.

[**Overall Summary**]

The experimental evaluation presented in this work has demonstrated that utilising this dialogue approach a 'correct' alignment can be generated between two ontologies, assigned to individual agents. A 'correct' mapping included in an alignment is one which is found by the approach which also features in a benchmark alignment, when utilising ontologies taken from the OAEI and the corresponding reference alignments. These alignments were successfully generated across the ontology pairs within the bounds of the neighbourhood threshold used within this evaluation. The results from the experimental evaluation of the three variants of this approach are summarised in Table 8.12 which illustrates the differences of the results across the variants.

The summary presents the mapping restrictions placed on the approach version, the sharing strategy for the testify move as a batch or single strategy, the neighbourhood threshold values where an alignment can be generated across all the datatypes evaluated, and finally neighbourhood threshold bound where the precision and recall converge illustrating the 'best values' for the metrics over each of the dialogue versions.

DbMN	**Mapping Restriction**	**Sharing Type**	**Alignment range (lowest bound)**	**Range (highest bound)**	**Precision, Recall Convergence**
_5	1:*	single	0	0.650	0.525-0.600
_6	1:*	batch	0	0.650	0.525-600
_7	1:1	single	0	0.750	0.450-0.550
DbMN	**Average** sharing $\epsilon_n = 0$	**Unshared** $\epsilon_n = 0...1$	**Sharing at** $\epsilon_n = 1$	**% Shared** $\epsilon_n = 0$	**% Shared** $\epsilon_n = 1$
_5	34.7	32.2	2.5	55.3	48.1
_6	35.4	32.2	3.2	56.5	48.1
_7	48.1	32.4	15.7	76.8	48.1

TABLE 8.12: Experimental evaluation summary of all three DbMN versions and average results when compared to a reference alignment

8.4 DbMN compared to current alignment approaches.

In this section, the performance of the two variants DbMN_6 and DbMN_7 are compared with a selection of other ontology matching systems, (discussed in Section 3.3.3). These systems include:

AML	AOT	AOTL	CroMatcher
DKPAOM	Gmap	JarvisOM	Lily
LogMap	LogMapC	LogMapLite	MassMatch
Mamba	OMReasoner	RESLWB	Xmap

These systems were selected as they provided alignments generated for the ontology pairs used in this evaluation, which are part of the OAEI dataset.

The selected current alignment systems were evaluated in comparison to the ontology pairs used in the DbMN approach detailed in Chapter 7, using the precision and recall measures, over a benchmark standard developed for this evaluation. Figure 8.18 and Figure 8.20 present the results of DbMN_5 and DbMN_6 in comparison to the other alignment systems. Figure 8.19 and Figure 8.21 present the results of the DbMN_7 variant.

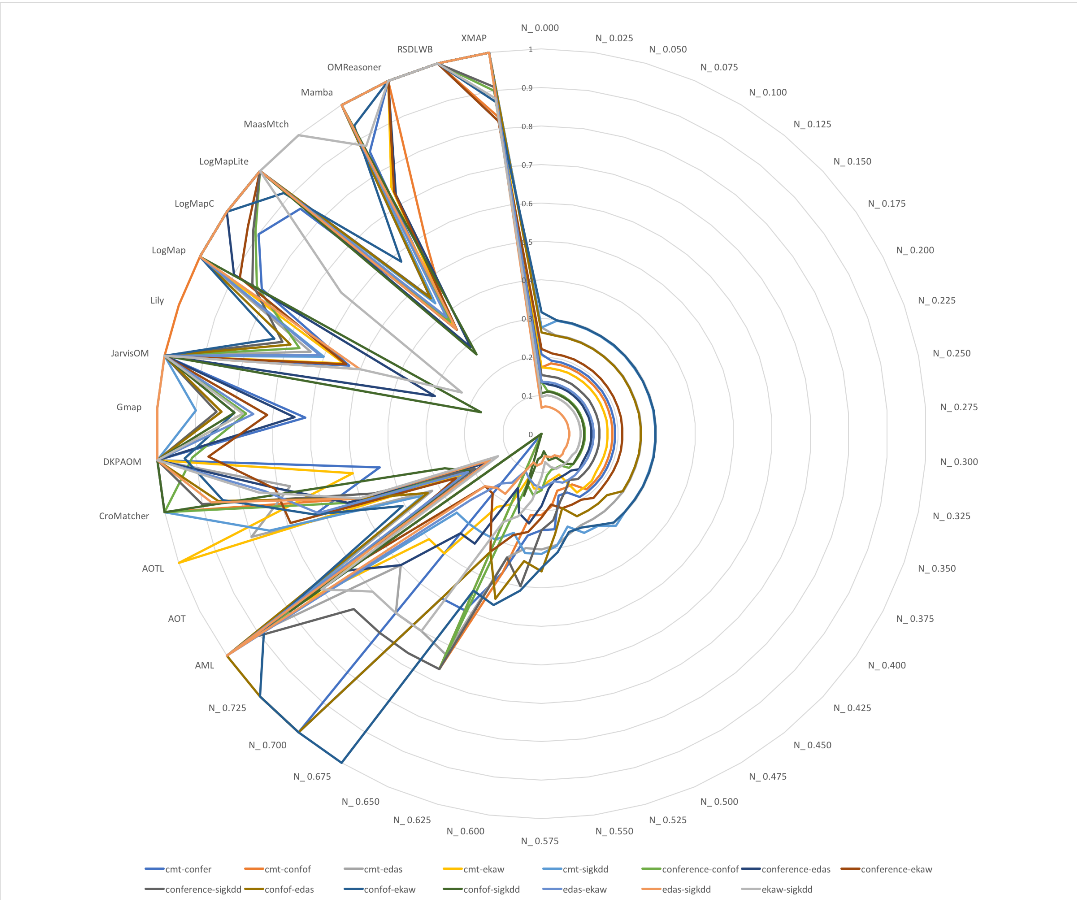


FIGURE 8.18: DbMN_5 and DbMN_6 for precision compared to current systems

Precision values of the DbMN approach compared with other alignment systems

Over all the ontologies pairs which were evaluated using the experimentation presented in this chapter, DbMN_5 and DbMN_6, yielded the same results, however only DbMN_6

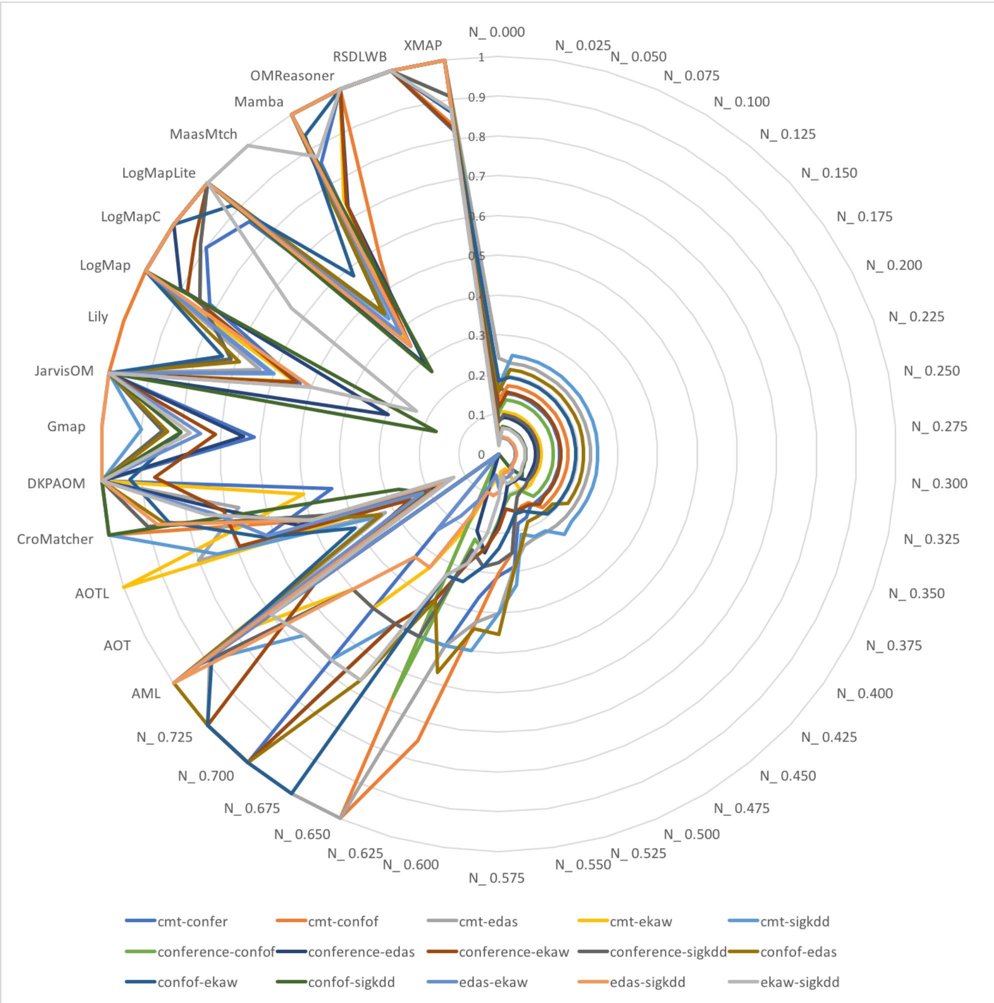


FIGURE 8.19: DbMN_7 for precision compared to current systems

is included in this evaluation section comparing to exsisting approaches. Using the precision results shown in Figure 8.18 as a way to address soundness, it can be seen that the precision for DbMN_6 is low when compared to the other systems with values of ϵ_n=[0..0.525], however is more comparable to current systems between the values of ϵ_n=[0.675..0.725].

Similarly for DbMN_7 (Figure 8.19) the precision value is also low when the neighbourhood threshold was in the range ϵ_n=[0..0.600]. Over the majority of ontology pairs evaluated the DbMN approach performs better given the precision results in comparison to the Lily and MaasMatch systems, however the remaining current systems perform better or match the results regarding the precision values found using the DbMN approach.

Recall values of the DbMN approach compared with other alignment systems

Using the recall results illustrated in Figure 8.20 to address completeness, it can be seen that both of the DbMN_5 and DbMN_6 variants produce consistently high recall in comparison to the other systems for the thresholds in the range ϵ_n= [0..0.425] and for DbMN_7 (Figure 8.21) in the range ϵ_n=[0..0.400]. This illustrates that the DbMN approach produces comparable recall results to those established by current systems. This could also be due to the DbMN producing larger alignments which include more mappings (both correct and incorrect as discussed in Section 8.3).

Between the threshold values of ϵ_n= [0.425..0.725] the DbMN recall values decrease and the current systems on average, perform better in terms of completeness, this can be explained in the DbMN approach by high neighbourhood threshold overfitting the candidate correspondences.

8.5 DbMN Approach Conclusions

The dialogue based approach has illustrated the ability to generate meaningful alignments over real world ontologies. Using the results for the three variants presented in this chapter, the approach found meaningful alignments over the ontologies between an incremented bound of the increased neighbourhood values. Within this evaluation, the following variables were investigated in order to see their influence on the generated alignment resulting from the DbMN approach.

[V.1] The lexical similarity threshold over these experiments was kept at 0 in order to eliminate the influence of the neighbourhood similarity metric, and the influence of the structural similarity in generating an alignment between the two inputted ontologies.

[V.2] The neighbourhood similarity threshold was incremented for each of the variants by a value 0.025 within the range [0..1]. In the evaluation, across all the variants it was shown that this threshold, when increased, removed the *incorrect* mappings and gradually improved the resulting alignment. As the threshold increase, a point was reached where the metric overfitted the correspondences and the approach could no longer find any accepted mappings to be included in the resulting alignment generated. Although this was found at different points between different variants, the overall trend was consistent suggesting that there was an optimum bound between 0 and 1, where the approach found the best aggregate alignment across the precision and recall (with respect to the benchmark alignments).

[V.3] The influence on the order of concepts in the signature $\boldsymbol{\Sigma}^t$ was found to only affect the variant with the 1:1 mapping. This was due to the concept which was shared first by the opponent agent, in response to the concept to be mapped from the signature, being be mapped and thus not being considered as a candidate for other mappings. The ordering of the initiated concepts thus determined which concepts were mapped. This however did not have an effect on the alignment in terms of mappings found in the benchmark, as these were still accepted, on a similar scale to when this was not utilised as a variable as in DbMN_5 and 6.

[V.4] This variable investigated the ranking function or a random order in which the triples were in the formation of the premise to support the candidacy of a mapping. This

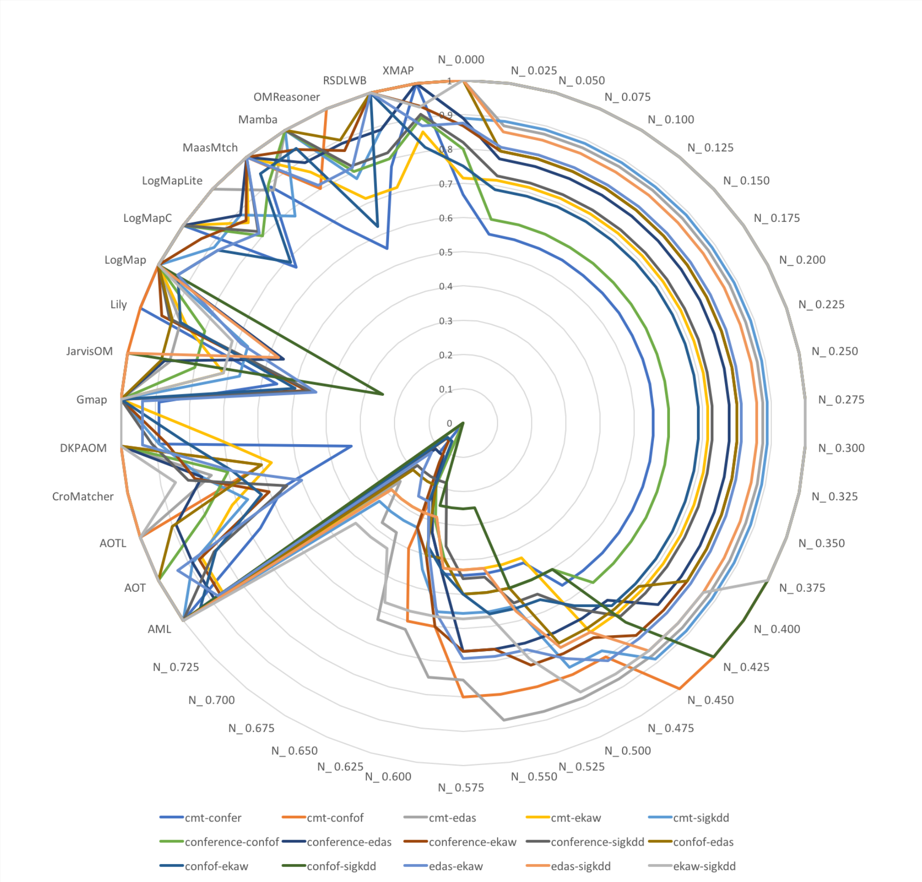


FIGURE 8.20: DbMN_5 and DbMN_6 for recall compared to current systems

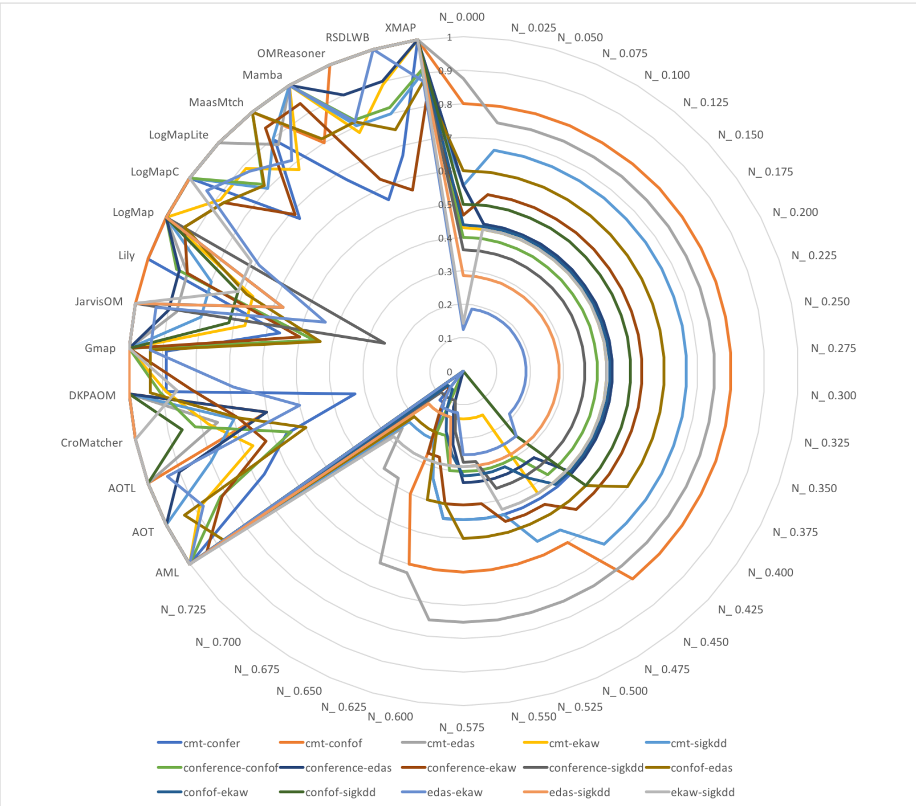


FIGURE 8.21: DbMN_7 for recall compared to current systems

was found not to have an effect on the approach, as with both variables, this generated the same results with respect to the resulting alignments.

The DbMN approach, evaluated across all of these settings, was found to be successful when the neighbourhood threshold was in the range [0..1], in generating a meaningful alignment supported with a premise providing structural support to the correspondences. However, its performance when directly compared to current systems, although adopting a different strategy is poor. For the task of maintaining a level of privacy for the agents, in particular the opponent agent, the system is able to maintain meaningful outputs, and provide the agent with a significant level of privacy, in terms of concepts that remain unshared, throughout the process.

The first dialogue variant, where sharing the neighbourhood was done in a single full move (using a batch approach), provided the agents, with the same information as when sharing is done individually. However, by sharing all of the triples in one move, the agents were able to find acceptable mappings more rapidly whilst maintaining a level of privacy. The second dialogue variant used in the evaluation evolved from generating one to many mappings in an alignment (in contrast to generating one to one mappings). This restriction, allowed the DbMN approach to generate a more accurate alignment pruning out *incorrect* mappings.

The 1:1 mapping restriction provides strong results, in that fewer *incorrect* mappings were found. However, this is influenced by the ordering of the signature to be mapped. As a result, an improvement of the approach would be to generate a more pertinent and meaningful ranking method for ordering the concepts in the proponent's signature, in terms of the concepts that are more important for the agent to be mapped correctly.

From these results, the hypotheses were supported across all the variants of the approach. As the neighbourhood value increased toward the value of 1, each variant pruned out *incorrect* mappings until they overfitted these candidate mappings and no alignments could be found.

The experiments designed for this evaluation of the DbMN approach did not aim to find the best values, in terms of configuring the protocol to produce the best results when evaluated against a benchmark. However whilst some considerations in terms of the neighbourhood threshold value for higher recall results can be made, these can not be generalised as configuration settings and are the subject of future work.

Overall, DbMN addressed the research question in that a meaningful alignment could be found utilising this approach of ontology matching, and providing a level of privacy for the agents. There is potential in further evaluate this approach in order to establish how the optimum values of these variables, is the alignment generation. Some of these avenues of further work are outlined in Chapter 9.

8.6 Summary

This chapter details how the protocol has been used in order to evaluate this dialogue based approach to ontology mapping. This chapter has presented the evaluation of three defined versions of the DbMN approach, detailing their differences and individual restrictions in comparison to each other, in order to evaluate the performance in generating meaningful alignments to answer the defined research questions. The variants have been empirically evaluated using precision and recall compared to a benchmark standard, and finally an overview of the DbMN performance in comparison to selected current systems is summarised.

In Summary of this Chapter:

- Detailed the experiment preliminaries that hold across the three different evaluated versions of the DbMN approach.
- Detailed the evaluation of DbMN_5, and a discussion of the outputted alignments results generated by the system, empirically evaluated using precision and recall.
- Presented an evaluation of the second DbMN_6, differing from the first in the sharing strategy adopted by the agents. These results were then presented and discussed in comparison to the previous version.
- Finally the third version of the approach has been presented and discussed in comparison to the two previous versions, and empirically evaluated.
- This chapter presents the results of the DbMN approach in comparison to current ontology matching systems, in terms of precision and recall to a developed bench standard.
- Finally this chapter discussed what was found from the evaluation of the approach, and proposes further experimentation to develop it.

Part IV

Synopsis

Chapter 9

Conclusions and Future Work

Chapter Outline

'The end of a melody is not its goal: but nonetheless, had the melody not reached its end it would not have reached its goal'. - Friedrich Nietzsche

This chapter marks the final of this thesis and presents a summary of the research that has been conducted along with potential avenues for further work. Firstly Section 9.1 presents a review of the contribution of the novel dialogue based approach to ontology matching and the implications of this work are discussed, alongside its main research achievements. The Section 9.2 presents the potential avenues of future work, which have been identified in order to further this contribution.

9.1 Review of Contribution

As stated in the introduction of this thesis, this research sets out to answer the main research question of: *'Can a meaningful alignment be generated over two ontologies, when knowledge is not shared a priori?'*. In order to answer this research question, the following goals were put forward:

i To develop a decentralised ontology matching system based on a dialogical protocol, allowing two participating agents to attempt to generate a meaningful alignment between their knowledge bases. Where a meaningful alignment is supported by semantic and syntactic knowledge enhancing the meaning of the correspondence found;

ii To allow the participating agents to utilise this designed protocol and generate a meaningful alignment without the agents having to share all of the knowledge within their ontology to either.

The work presented in this thesis has addressed this research question and developed a working dialogue protocol and accompanying agent strategy. This research was successful in reaching its goals of developing a dialogical system in order to generate meaningful ontology alignments in a decentralised approach using negotiation, where participating agents negotiate over meaning in an attempt to generate a meaningful alignment, by incrementally sharing only parts of their full ontology. In this section, a summary of the work undertaken in answering the above research question is presented, along with the contribution and how the defined research goals were achieved. The research aims are detailed specifically in Chapter 1.

Chapter 1 introduced the research area including an overview of the areas of ontologies and ontology alignment in relation to the objectives of the research.

In Chapters 2, 3 and 4 a literature review is presented where these research areas were further discussed and detailed. This began with Chapter 2 which covered a detailed discussion of definition of ontologies, their conceptualisation and categorisation, and the main components that generate them, followed by an overview of ontology languages. Chapter 3 developed the use of ontologies in this work and presented an introduction into the ontology alignment generation. Ontology alignment generation was the key in the contribution of this work as a dialogue is presented to address this specific problem. This ontology alignment problem was defined in terms of the heterogeneity that occurs due to the subjective and individual design of ontologies, specific to particular services and designed to address defined task. This chapter provided detailed examples of the different types of heterogeneity and detailed how this problem is addressed with reference to current alignment systems.

In Chapter 4, the concept of dialogues was introduced, documenting the importance of communication between agents in attempting to address the previously defined ontology alignment problem. The importance of this communication was approached from philosophical and psychological background using human conversation and speech acts as a model to provide agents with a particular set of statements available in an interaction. This chapter introduces the concept of a dialogue and how agents interact in order to achieve a goal and it also provides a setting for casting the ontology alignment problem as a dialogue game. Here, the concept of using autonomous agents within a dialogue setting was introduced and the notion of speech acts as available locutions within this game were put forward.

Chapter 5 followed the literature review and presented the main introduction of the contribution, and detailed one of the main goals of the research to develop a decentralised ontology matching system, based on a dialogical protocol, which allowed two participating agents to attempt to generate a meaningful alignment between their knowledge bases, supported by semantic and structural knowledge enhancing the meaning of the correspondence found. This protocol was detailed in terms of the available moves the agents could utilise and which had been developed from speech acts detailed in Chapter 4, and in line with the terms of conversational rules. The formalised moves were followed with the definition of the decision mechanisms used within this system in allowing an alignment to be accepted as meaningful by the participating agents. This chapter continued with the termination proof of the dialogue, and followed with the detailing of the strategy used within this protocol by the agents in order to travers the dialogue and accept correspondences mutually through negotiation.

Chapter 6 presented a detailed walkthrough example of the dialogue protocol presented in the previous chapter. This example provided a step by step presentation of how the dialogue is used by two participating agents, *Alice* and *Bob* in aligning their two defined ontology fragments.

Chapter 7 introduced the explicit parameters used in this dialogue for the evaluation of the DbMN approach, and detailed the methodology used. This included the parameters to which the experiments have been designed. These included a justification of the lexical similarity method selected for this approach, and also defined the specific rank function assumed by the agents.

In Chapter 8, the dialogue protocol and the agents' strategy were evaluated using an experimentation in addressing the ontology alignment problem and its variants, in its ability to address the research question posed in Chapter 1. This implementation presented an overview of the specific mechanisms used in the experimentation and the specific agent strategies adopted for the analysis of the approach. The experiments presented the results in an empirical evaluation of the approach evaluated against the variants and current alignment systems used in the OAEI. This chapter finally presented an overview of the further experimentation that could be utilised in order to further develop and evaluate the performance of the approach.

This final contribution of the thesis presents a working decentralised approach which was developed in order to address the research hypothesis. The dialogue approach allowed two participating agents to negotiate corresponding mappings and mutually accept an alignment in a decentralised environment, whilst not having to fully disclose their knowledge bases to each other or to a third party.

9.2 Future Work

During the course of this research, a variety of areas of interest became apparent which due to time constraints could not be further developed within this work. This section summarises these areas which are thought to have noticeable potential for future research and outline a possible direction for furthering the ideas presented in this thesis.

Future work could include a reversal of the ontologies assigned within the experiments conducted and presented in Chapter 8 in order to evaluate if the alignments generated are symmetrical, and how the sharing differs from one assignment to another. Table 9.1 illustrates the reversed assignments of the ontologies used within the evaluation of Chapter 8. The experimentation could be repeated to include these ontology pairs and the resulting alignments compared to those results presented in Chapter 8.

TABLE 9.1: Ontology Pairs reversed

Proponent	**Opponent**	**Proponent**	**Opponent**
conference	cmt	edas	conference
confof	cmt	sigkdd	conference
sigkdd	cmt	ekaw	conference
ekaw	cmt	confof	conference
edas	cmt	ekaw	edas
sigkdd	confof	sigkdd	edas
ekaw	confof	sigkdd	ekaw
edas	confof	-	-

As discussed in Chapter 8, the evaluation of the DbMN approach was conducted using real world ontologies taken from the OAEI conference dataset. These ontologies however are all of a similar size and can be categorised as small ontologies in comparison to the large biomedical datasets available on the OAEI. This biomedical track consists of three ontologies Foundational Model of Anatomy (FMA) [100], SNOMED CT [37] and National Cancer Institute Thesaurus (NCI) [95] all of which are very semantically rich, and contain a minimum of ten thousand classes, and potentially represent entities with more than one label. One avenue of future work could be to use large more semantically rich ontologies, such as those presented in the biomedical track to evaluate the DbMN approach in generating alignments over a much large dataset. This line of evaluation would be beneficial to investigate the running time of the DbMN approach and potentially evolve the approach to be less computational expensive.

The dialogue presented uses *internal* matching techniques (string based methods), to establish similarity between the term labels in the ontologies. In contrast to this *external* matching techniques can utilise background knowledge in order to 'bridge the semantic gap' [110] to establish meaning. An alternative direction for furthering the DbMN approach could be to adjust the string based similarity methods used to establish similarity between concepts, to utilise alternative semantic frameworks. Background knowledge has been used within the ontology matching community [6, 53, 110] to provide methods of generating alignments by taking advantage of available resources in establishing meaningful matches. Context-based ontology matching [83] uses intermediate resources available to agents to which their ontologies can be connected.

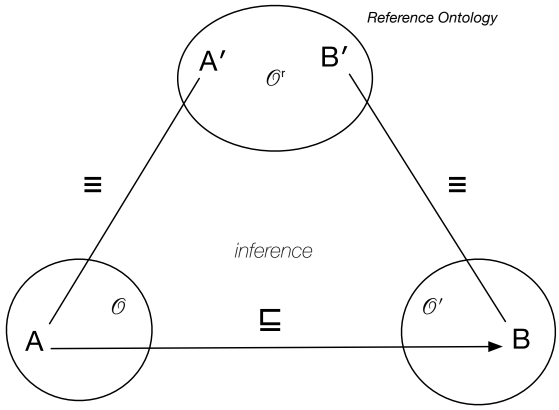


FIGURE 9.1: Exploiting background knowledge to establish meaning using anchor concepts (A', B'), to infer a meaningful match between A and B [110]

This context-based approach (seen in Figure 9.1) illustrates the use of background knowledge in establishing an 'anchor' concept in order to provide semantic meaning to infer a match between the concepts $A \in O$ and $B \in O'$. This context-based approach could exploit explicit relations between the agents ontologies, and intermediate resources in order to provide evidence to establish similarities between concepts between the participating agents assumed ontologies. Semantic resources such as WordNet [46], which is developed as a semantic dictionary, could also provide the agents with background knowledge as synonyms of terms, which they could utilise to establish further meaning, in order to find a more accurate mapping. Utilising a context-based approach could also be used in order to further this work to address mappings where ontologies are represented in different natural languages, by using language translation resources such as BabelNet [96]. Such resources provide multilingual background knowledge which is needed to identify different language labels referring to the same concept.

Future work could also include adding further agents into the dialogue as illustrated in Figure 9.2. The inclusion of more agents however, would require further restrictions and conditions added to both the pre and post conditions of each of the locutions, in order address turn taking and to allow all the agents to accurately and individually exchange information in order to generate their required alignments.

The dialogue approach presented upholds a restriction on retraction of commitment, it is assumed that once a statement is made, the agents are permanently committed to it. This prevents the agents from revisiting accepted mappings, once they are committed in *CS* thus resulting in the alignments being order dependant. Relaxing this restriction and permitting the agents to revise previously committed mappings, would allow the agents the ability to propose a counter argument or *rebuttal*, if it was established that

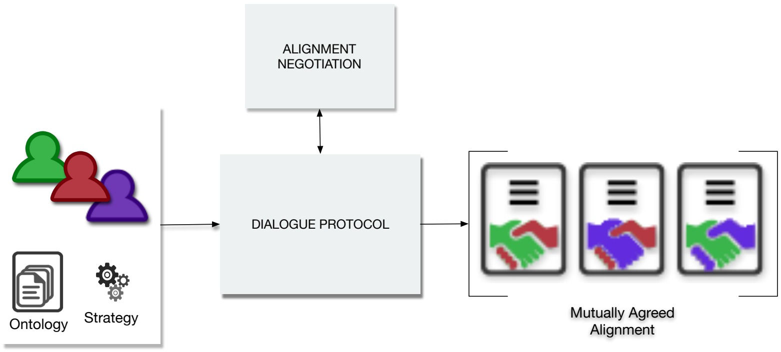


FIGURE 9.2: Conceptual architecture for multi-agent approach.

throughout the dialogue further knowledge is shared to better support an alternative mapping. This retraction could be permitted through the use of a new locution to establish this counter argument, or potentially through re-adjusting pre and post conditions of the current locutions available to the agent. This retraction would have to be developed taking into consideration the dialogue termination, such that if a rebuttal is posed against a given argument, the agents will still mutually accept or reject candidate mappings such that cycles of arguments do not occur, thus ensuring the dialogue will terminate.

Another element that could be included in future work is that of how the commitment store and their contents are removed following the end of a dialogue between two participating agents. This could be investigated in order to maintain a store of previous alignments, in order to potentially reduce the negotiation time for mappings. If the agents have a store of previous alignments, they could use this as a premise (PR) within a dialogue to generate an alignment more quickly within this new dialogue run. This would require some adjustments to be made on the pre and post conditions of the locutions, in order to ensure termination.

A current constraints of the existing approach is that neighbourhood size is only considered with one edge beyond the concept under consideration. Further work could relax this restriction and investigate how a larger neighbourhood of up to n edges could be explored, illustrated in Figure 9.3.

This further work would need to take into account possible redundancy in this exploration due to the fact that neighbourhoods will overlap, and the inherent potential exponential increase in complexity in determining a neighbourhood match. This is exacerbated by the fact that a neighbourhood similarity is determined through the evaluation of different matchings as detailed in Chapter 5.

For the purposes of implementing this DbMN dialogue assumptions and restrictions were made on this approach regarding strategic decision making as detailed in Chapter 7. Future work investigating this DbMN approach could relax and adapt these restrictions and assumptions to adjust the following:

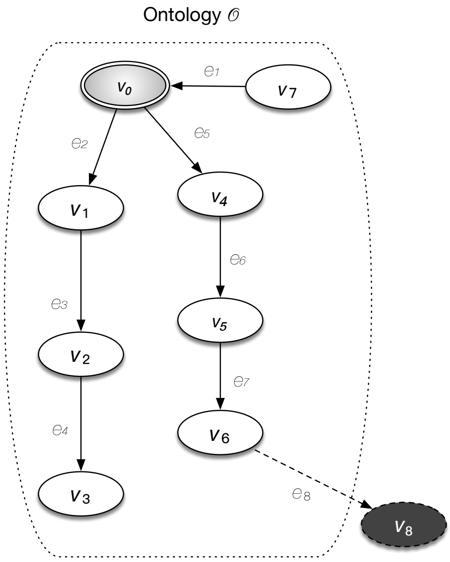


FIGURE 9.3: Exploring a larger neighbourhood.

Lexical matching method: This approach was implemented with the Jaro-Winker lexical matching method as a string comparator for the lexical similarity metric. A possible direction of future research could use a variation of different lexical matching techniques such as those presented in Chapter 7. Varying the choice of the lexical matching technique or adding further methods to the existing string based matcher selected, could demonstrate different entities selected as similar due to the string based comparisons used in the methods.

Neighbourhood threshold: Chapter 7 presents the incremented step value of the neighbourhood threshold used within this approach, which has been utilised within the implementation. This threshold was set to increment by steps of 0.025 between [0..1]. A more comprehensive step value of this neighbourhood threshold could be used in order to present the differences between each alignment in more detail.

Signature Σ^t contents: Relaxing the current restriction on the approach to include the full proponents disclosable ontology in the signature Σ^t. Including only a subset of the proponent's ontology in this signature, would potentially allow the proponent agent a degree of privacy in the sharing of their assigned ontology throughout the dialogue.

Alternate combination of variables: The implementation of the DbMN dialogue consisted of three variations of approach restricting the generated alignments to 1*1 mappings and how the agents shared the neighbourhood of the concept under negotiation at the *testify* move. A fourth variant using these variables could be to combine these two restrictions to further evaluate the approach.

Lexical threshold: As detailed in Chapters 7 and 8 the DbMN uses both a lexical similarity and a structural similarity in generating an alignment. The restrictions on the implemented version of this approach, isolated the structural similarity in order to evaluate the performance of the approach where the lexical threshold was set to 0. Relaxing this restriction would allow an evaluation in the influence of both the lexical and structural thresholds on the alignments generated. These generated alignments could be compared with the alignment results presented in this research.

Although the approach presented has successfully answered the research questions posed and has met the objective of establishing a mutually agreed alignment, there are assumptions limiting this approach. Relaxing these assumptions and adapting the approach as detailed in this chapter provides an obvious avenue of further work to develop this dialogue to further investigate the DbMN capabilities in generating alignments whilst maintaining agent's privacy.

Part V

Appendices

Appendix A

DbMN Experimentation Results

Chapter Outline

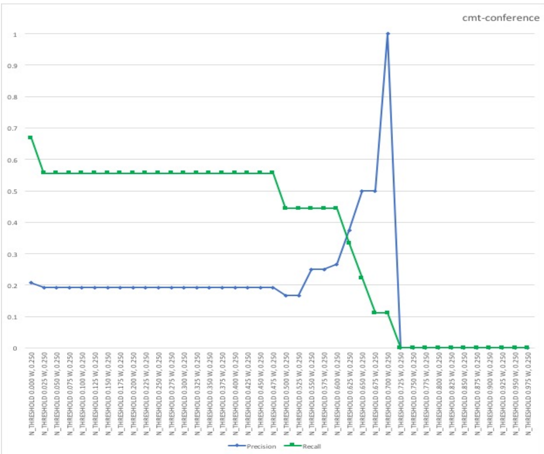
This Appendix (A) documents further results taken from the experiments conducted over the dialogue approach variants DbMN_5, DbMN_6 and DbMN_7 described in Chapter 8 of this thesis. In contrast to the graphs detailed and discussed in Chapter 8, the graphs in this section represent each of the ontology dataset pairs individually including the precision and recall values, and the number of mappings and correct mappings each approach variant finds. This appendix also illustrates the number and percentage of the opponent's ontology (O') shared throughout the three variants of the approach.

A.1 DbMN_5 Experimentation Results

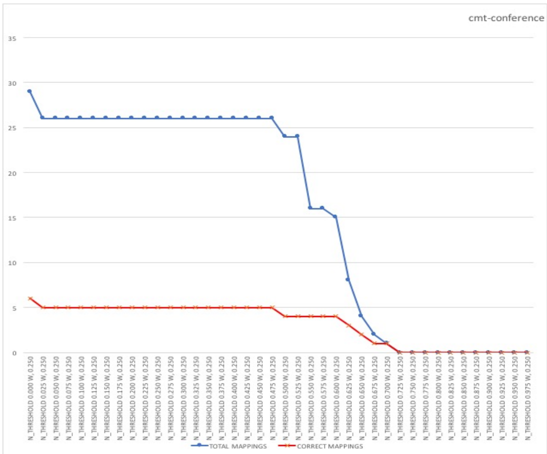
This appendix documents the results for the individual experiments for both the DbMN_5 and DbMN_6 versions of the approach evaluated. These two versions DbMN_5 and DbMN_6, produce the same outputted results as discussed in Chapter 8 so have been included here as a single set of results. These results have been discussed in Chapter 8 and are included in this appendix to illustrate each of the results for the dataset pairs individually, to accompany the combined results presented in that chapter.

The results as presented here, show two graphs for each dataset pairs, as detailed in Chapter 8, where the x axis shows the incremented neighbourhood threshold values, and the y axis shows the precision and recall values for the left graph, and the number of mappings on the right. The left graph of each pair, presents the precision and recall values between ϵ_n=[0..1], and the right presents the number of mappings found by the approach and the number of 'correct mappings' found which are included in the platinum standard benchmark.

R.1 **For the ontologies O = cmt, $O\prime$ = conference:**



(a) O = cmt, $O\prime$ = conference Precision and Recall



(b) O = cmt, $O\prime$ = conference Total Mappings

FIGURE A.1: Results for O = cmt, $O\prime$ = conference

Graphs illustrating the precision and recall values for O = cmt, $O\prime$ = conference, with the highest precision = 1 and the highest recall = 0.67. The graph on the right illustrates the total mappings found from 29 to 0, and the mappings found by the system that feature in the *platinum standard* i.e. 'correct mappings' between O = cmt, $O\prime$ = conference, from the values 6 mappings to 0.

R.2 **For the ontologies O = cmt, $O\prime$ = confof:**

Graphs illustrating the precision and recall values for the ontologies O = cmt, $O\prime$ = confof, with the highest precision = 0.172 and the highest recall = 1. The graph on the right illustrates the total number of mappings found from 29 to 0, and the

(a) O = cmt, O' = confof Precision and Recall

(b) O = cmt, O' = confof Total Mappings

FIGURE A.2: Results for O = cmt, O' = confof

mappings found by the system that feature in the *platinum standard* from the values 5 mappings to 0.

R.3 **For the ontologies O = cmt, O' = edas:**

Graphs illustrating the precision and recall values for the dataset pair O = cmt, O' = edas, with the highest precision = 0.276 and the highest recall = 1 presented on the left. The graph on the right illustrates the total number of mappings found from 29 to 0, and the mappings found by the system that feature in the *platinum standard* benchmark from the values 8 mappings to 0.

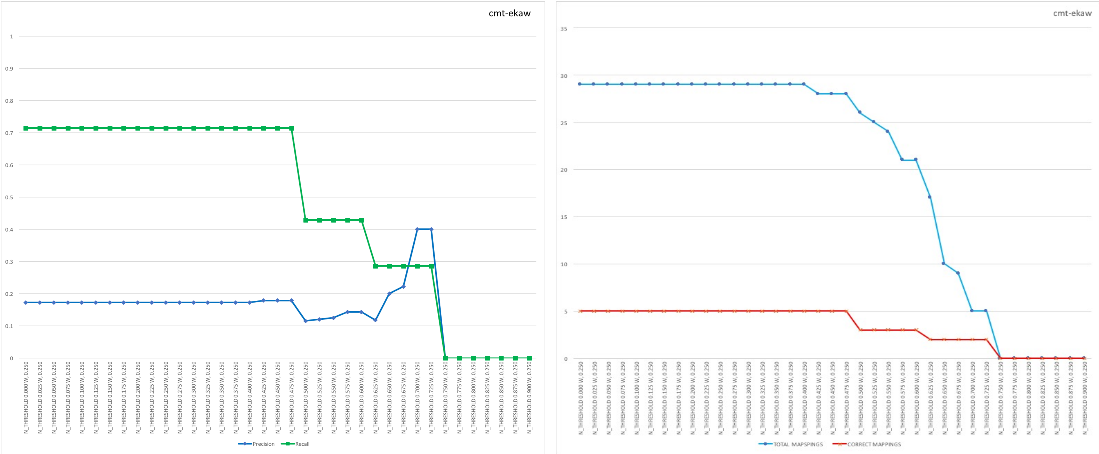
(a) O = cmt, O' = edas Precision and Recall

(b) O = cmt, O' = edas Total Mappings

FIGURE A.3: Results for O = cmt, O' = edas

R.4 **For the ontologies O = cmt, O' = ekaw:**

Graphs illustrating the precision and recall values (left) for the ontologies O = cmt, O' = ekaw, with the highest precision = 0.172 and the highest recall = 0.714. Secondly the total mappings found (right) from 29 to 0 mappings, and the number of mappings found by the system that feature in the *platinum standard* from the values 5 mappings to 0.



(a) O = cmt, $O\prime$ = ekaw Precision and Recall

(b) O = cmt, $O\prime$ = ekaw Total Mappings

FIGURE A.4: Results for O = cmt, $O\prime$ = ekaw

R.5 **For the ontologies O = cmt, $O\prime$ = sigkdd:**

Graphs illustrating the precision and recall values for O = cmt, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.276 and the highest recall = 0.889. Secondly on the right the total mappings found from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, $O\prime$ = sigkdd, from the values 8 mappings to 0.

(a) O = cmt, $O\prime$ = sigkdd Precision and Recall

(b) O = cmt, $O\prime$ = sigkdd Total Mappings

FIGURE A.5: Results for O = cmt, $O\prime$ = sigkdd

R.6 **For the ontologies O = conference, $O\prime$ = confof:**

Graphs illustrating the precision and recall values for O = conference, $O\prime$ = confof from ϵ_n=[0..1], with the highest precision = 0.136 and the highest recall = 0.8. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, $O\prime$ = confof, from the values 8 mappings to 0.

(a) O = conference, $O\prime$ = confof Precision and Recall (b) O = conference, $O\prime$ = confof Total Mappings

FIGURE A.6: Results for O = conference, $O\prime$ = confof

R.7 **For the ontologies O = conference, $O\prime$ = edas:**

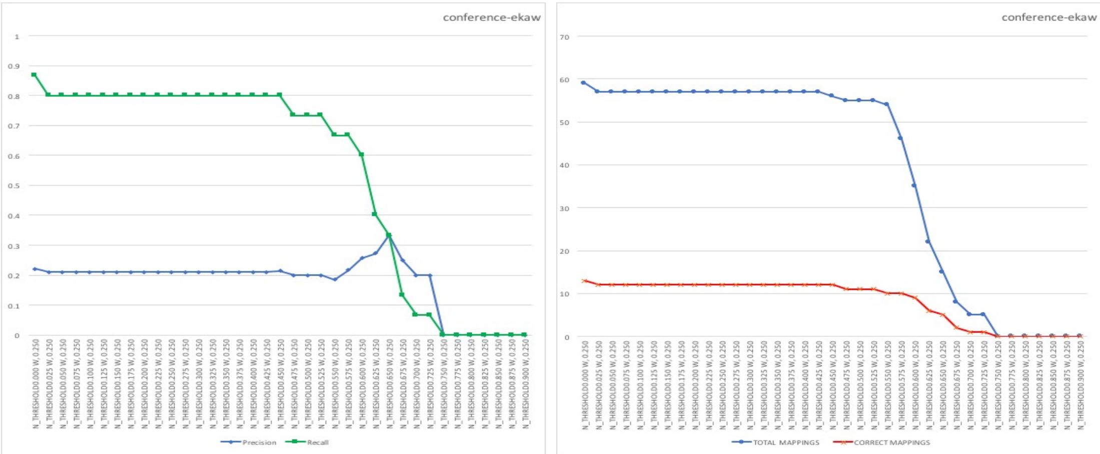
Graphs illustrating the precision and recall values for O = conference, $O\prime$ = edas from ϵ_n=[0..1], with the highest precision = 0.136 and the highest recall = 0.889. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, $O\prime$ = edas, from the values 8 mappings to 0.

(a) O = conference, $O\prime$ = edas Precision and Recall (b) O = conference, $O\prime$ = edas Total Mappings

FIGURE A.7: Results for O = conference, $O\prime$ = edas

R.8 **For the ontologies O = conference, $O\prime$ = ekaw:**

Graph on the left illustrating the precision and recall values for O = conference, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 0.220 and the highest recall = 0.867. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = conference, $O\prime$ = ekaw, from the values 13 mappings to 0.



(a) O = conference, $O\prime$ = ekaw Precision and Recall (b) O = conference, $O\prime$ = ekaw Total Mappings

FIGURE A.8: Results for O = conference, $O\prime$ = ekaw

R.9 **For the ontologies O = conference, $O\prime$ = sigkdd:**

Graphs illustrating the precision and recall values for O = conference, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.153 and the highest recall = 0.812. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = conference, $O\prime$ = sigkdd, from the values 9 mappings to 0.

(a) O = conference, $O\prime$ = sigkdd Precision and Recall (b) O = conference, $O\prime$ = sigkdd Total Mappings

FIGURE A.9: Results for O = conference, $O\prime$ = sigkdd

R.10 **For the ontologies O = confof, $O\prime$ = edas:**

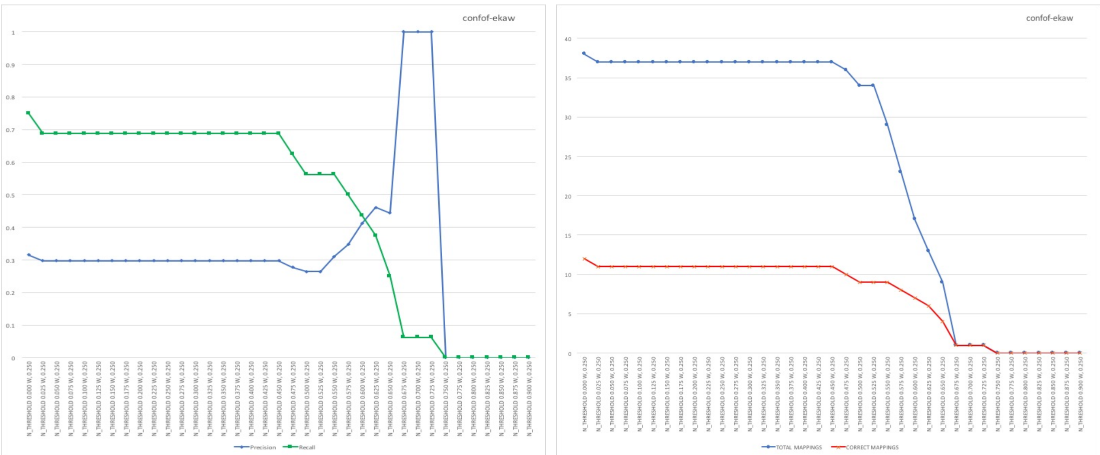
Graphs illustrating the precision and recall values for O = confof, $O\prime$ = edas from ϵ_n=[0..1], with the highest precision = 0.263 and the highest recall = 1. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = confof, $O\prime$ = edas, from the values 10 mappings to 0.

(a) O = confof, $O\prime$ = edas Precision and Recall (b) O = confof, $O\prime$ = edas Total Mappings

FIGURE A.10: Results for O = confof, $O\prime$ = edas

R.11 **For the ontologies O = confof, $O\prime$ = ekaw:**

Graphs illustrating the precision and recall values for$O$ = confof, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 0.316 and the highest recall = 0.75. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = confof, $O\prime$ = ekaw, from the values 12 mappings to 0.



(a) O = confof, $O\prime$ = ekaw Precision and Recall (b) O = confof, $O\prime$ = ekaw Total Mappings

FIGURE A.11: Results for O = confof, $O\prime$ = ekaw

R.12 **For the ontologies O = confof, $O\prime$ = sigkdd:**

Graphs illustrating the precision and recall values for O = confof, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.105 and the highest recall = 1. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = confof, $O\prime$ = sigkdd, from the values 4 mappings to 0.

(a) O = confof, $O\prime$ = sigkdd Precision and Recall (b) O = confof, $O\prime$ = sigkdd Total Mappings

FIGURE A.12: Results for O = confof, $O\prime$ = sigkdd

R.13 **For the ontologies O = edas, $O\prime$ = ekaw:**

Graphs illustrating the precision and recall values for O = edas, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 0.136 and the highest recall = 0.875. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = edas, $O\prime$ = ekaw, from the values 14 mappings to 0.

(a) O = edas, $O\prime$ = ekaw Precision and Recall (b) O = edas, $O\prime$ = ekaw Total Mappings

FIGURE A.13: Results for O = edas, $O\prime$ = ekaw

R.14 **For the ontologies O = edas, $O\prime$ = sigkdd:**

Graphs illustrating the precision and recall values for O = edas, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.068 and the highest recall = 1. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = edas, $O\prime$ = sigkdd, from the values 7 mappings to 0.

(a) O = edas, $O\prime$ = sigkdd Precision and Recall (b) O = edas, $O\prime$ = sigkdd Total Mappings

FIGURE A.14: Results for O = edas, $O\prime$ = sigkdd

R.15 **For the ontologies O = ekaw, $O\prime$ = sigkdd:**

Graphs illustrating the precision and recall values for O = ekaw, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.068 and the highest recall = 1. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. ‘correct mappings’ between O = ekaw, $O\prime$ = sigkdd, from the values 7 mappings to 0.

(a) O = ekaw, $O\prime$ = sigkdd Precision and Recall (b) O = ekaw, $O\prime$ = sigkdd Total Mappings

FIGURE A.15: Results for O = ekaw, $O\prime$ = sigkdd

A.1.1 Total mappings found by the DbMN_5 version of the approach:

The heatmap illustrated in Figure A.16 shows the decrease in the total mappings and the number of correct mappings found by the DbMN_5 version of the approach from the threshold value of $\sigma_n = 0.0 - 0.5$. Figure A.17 continues this from the threshold values of $\sigma_n = 0.5 - 1$.

	SYSTEM:	N_ 0.000	N_ 0.025	N_ 0.050	N_ 0.075	N_ 0.100	N_ 0.125	N_ 0.150	N_ 0.175	N_ 0.200	N_ 0.225	N_ 0.250	N_ 0.275	N_ 0.300	N_ 0.325	N_ 0.350	N_ 0.375	N_ 0.400	N_ 0.425	N_ 0.450	N_ 0.475	N_ 0.500
cmt-confof	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
cmt-conference	Total Maps	29	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	24
	Correct Maps	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4
cmt-edas	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
cmt-ekaw	Total Maps	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	28	28	28	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3
cmt-sigkdd	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	25	25
	Correct Maps	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7
conference-confof	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	52	52
	Correct Maps	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
conference-edas	Total Maps	59	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	52	49	49	49
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6
conference-ekaw	Total Maps	59	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	56	55	55
	Correct Maps	13	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	11
conference-sigkdd	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	50	38
	Correct Maps	9	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	6
confof-edas	Total Maps	38	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	30	30	30	30
	Correct Maps	10	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7	7	7
confof-ekaw	Total Maps	38	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	36	34
	Correct Maps	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	10	9
confof-sigkdd	Total Maps	38	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	35	30	28	28
	Correct Maps	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	2
edas-ekaw	Total Maps	103	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	94	94	87
	Correct Maps	14	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	12
edas-sigkdd	Total Maps	103	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	79	79	77	76
	Correct Maps	7	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
ekaw-sigkdd	Total Maps	73	69	69	69	69	69	69	69	69	69	69	69	69	69	69	69	62	61	61	61	61
	Correct Maps	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6	6	6

FIGURE A.16: Total mappings and correct mappings from $\sigma_n = 0.0 - 0.5$

	SYSTEM:	N_ 0.500	N_ 0.525	N_ 0.550	N_ 0.575	N_ 0.600	N_ 0.625	N_ 0.650	N_ 0.675	N_ 0.700	N_ 0.725	N_ 0.750	N_ 0.775	N_ 0.800	N_ 0.825	N_ 0.850	N_ 0.875	N_ 0.900	N_ 0.925	N_ 0.95	N_ 0.975	N_ 1
cmt-confof	Total Maps	26	26	21	19	14	8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cmt-conference	Total Maps	24	24	16	16	15	8	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	4	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
cmt-edas	Total Maps	27	26	24	20	20	15	8	5	5	4	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	7	7	6	6	5	5	3	3	2	0	0	0	0	0	0	0	0	0	0	0
cmt-ekaw	Total Maps	26	25	24	21	21	17	10	9	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
cmt-sigkdd	Total Maps	25	20	17	16	16	15	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	5	4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
conference-confof	Total Maps	52	52	46	34	25	22	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	5	5	5	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
conference-edas	Total Maps	49	48	42	32	17	14	13	3	3	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	6	6	4	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-ekaw	Total Maps	55	55	54	46	35	22	15	8	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	11	11	10	10	9	6	5	2	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-sigkdd	Total Maps	38	33	22	20	10	6	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	5	5	4	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-edas	Total Maps	30	25	19	14	12	9	6	4	2	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	4	4	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-ekaw	Total Maps	34	34	29	23	17	13	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	9	9	9	8	7	6	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0
confof-sigkdd	Total Maps	28	28	22	17	15	6	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
edas-ekaw	Total Maps	87	86	83	78	67	50	38	27	18	8	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	12	11	11	11	9	6	4	4	3	2	0	0	0	0	0	0	0	0	0	0	0
edas-sigkdd	Total Maps	76	65	51	39	36	26	19	11	11	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	4	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
ekaw-sigkdd	Total Maps	61	56	54	36	23	18	16	5	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	5	4	4	4	4	4	3	3	3	0	0	0	0	0	0	0	0	0	0	0

FIGURE A.17: Total mappings and correct mappings from $\sigma_n = 0.5 - 1$

A.1.2 Total mappings found by the DbMN_6 version of the approach:

The heatmap illustrated in Figure A.18 shows the decrease in the total mappings and the number of correct mappings found by the DbMN_6 version of the approach from the threshold value of $\sigma_n = 0.0 - 0.5$. Figure A.19 continues this from the threshold values of $\sigma_n = 0.5 - 1$.

	SYSTEM:	N_ 0.000	N_ 0.025	N_ 0.050	N_ 0.075	N_ 0.100	N_ 0.125	N_ 0.150	N_ 0.175	N_ 0.200	N_ 0.225	N_ 0.250	N_ 0.275	N_ 0.300	N_ 0.325	N_ 0.350	N_ 0.375	N_ 0.400	N_ 0.425	N_ 0.450	N_ 0.475	N_ 0.500
cmt-confof	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
cmt-conference	Total Maps	29	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	24
	Correct Maps	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4
cmt-edas	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
cmt-ekaw	Total Maps	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	28	28	28	26
	Correct Maps	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3
cmt-sigkdd	Total Maps	29	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	26	25	25
	Correct Maps	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7
conference-confof	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	52	52
	Correct Maps	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
conference-edas	Total Maps	59	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	54	52	49	49	49
	Correct Maps	8	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6
conference-ekaw	Total Maps	59	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	57	56	55	55
	Correct Maps	13	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	11	11
conference-sigkdd	Total Maps	59	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	50	38
	Correct Maps	9	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	6
confof-edas	Total Maps	38	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	30	30	30	30
	Correct Maps	10	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	7	7	7	7
confof-ekaw	Total Maps	38	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	36	34
	Correct Maps	12	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	10	9
confof-sigkdd	Total Maps	38	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	35	30	28	28
	Correct Maps	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	2	2
edas-ekaw	Total Maps	103	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	94	94	87
	Correct Maps	14	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	12
edas-sigkdd	Total Maps	103	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	83	79	79	77	76
	Correct Maps	7	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	5	5
ekaw-sigkdd	Total Maps	73	69	69	69	69	69	69	69	69	69	69	69	69	69	69	69	62	61	61	61	61
	Correct Maps	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	6	6	6	6

FIGURE A.18: Total mappings and correct mappings from $\sigma_n = 0.0{-}0.5$ using DbMN_6

	SYSTEM:	N_ 0.500	N_ 0.525	N_ 0.550	N_ 0.575	N_ 0.600	N_ 0.625	N_ 0.650	N_ 0.675	N_ 0.700	N_ 0.725	N_ 0.750	N_ 0.775	N_ 0.800	N_ 0.825	N_ 0.850	N_ 0.875	N_ 0.900	N_ 0.925	N_ 0.95	N_ 0.975	N_ 1
cmt-confof	Total Maps	26	26	21	19	14	8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cmt-conference	Total Maps	24	24	16	16	15	8	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	4	4	4	4	4	3	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
cmt-edas	Total Maps	27	26	24	20	20	15	8	5	5	4	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	7	7	6	6	5	5	3	3	2	0	0	0	0	0	0	0	0	0	0	0
cmt-ekaw	Total Maps	26	25	24	21	21	17	10	9	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
cmt-sigkdd	Total Maps	25	20	17	16	16	15	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	5	4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
conference-confof	Total Maps	52	52	46	34	25	22	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	5	5	5	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
conference-edas	Total Maps	49	48	42	32	17	14	13	3	3	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	6	6	4	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-ekaw	Total Maps	55	55	54	46	35	22	15	8	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	11	11	10	10	9	6	5	2	1	1	0	0	0	0	0	0	0	0	0	0	0
conference-sigkdd	Total Maps	38	33	22	20	10	6	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	6	5	5	4	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-edas	Total Maps	30	25	19	14	12	9	6	4	2	2	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	7	5	5	5	4	4	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
confof-ekaw	Total Maps	34	34	29	23	17	13	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	9	9	9	8	7	6	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0
confof-sigkdd	Total Maps	28	28	22	17	15	6	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
edas-ekaw	Total Maps	87	86	83	78	67	50	38	27	18	8	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	12	11	11	11	9	6	4	4	3	2	0	0	0	0	0	0	0	0	0	0	0
edas-sigkdd	Total Maps	76	65	51	39	36	26	19	11	11	10	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	5	4	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
ekaw-sigkdd	Total Maps	61	56	54	36	23	18	16	5	5	5	0	0	0	0	0	0	0	0	0	0	0
	Correct Maps	6	5	4	4	4	4	4	3	3	3	0	0	0	0	0	0	0	0	0	0	0

FIGURE A.19: Total mappings and correct mappings from $\sigma_n = 0.5 - 1$ using DbMN_6

A.1.3 Comparing DbMN_5 to other Ontology Matching systems

The DbMN_5 version of the approach has been compared to other ontology matching systems evaluated in the OAEI [2]. The approach has been compared using precision and recall values compared to a benchmark standard.

Precision:

The Graph A.20 shows the results of the precision from all the pairs of ontologies, using the DbMN_5 version of the approach. The precision is found over all the incriminations of the N_threshold value compared against some of those found by the systems discussed in Chapter 3 including: AML, AOT, AOTL, CROMatcher, DKPAOM, GMap, JarvisOM, Lily, MassMatch, Mamba, OMReasoner, RSDLWB and XMAP.

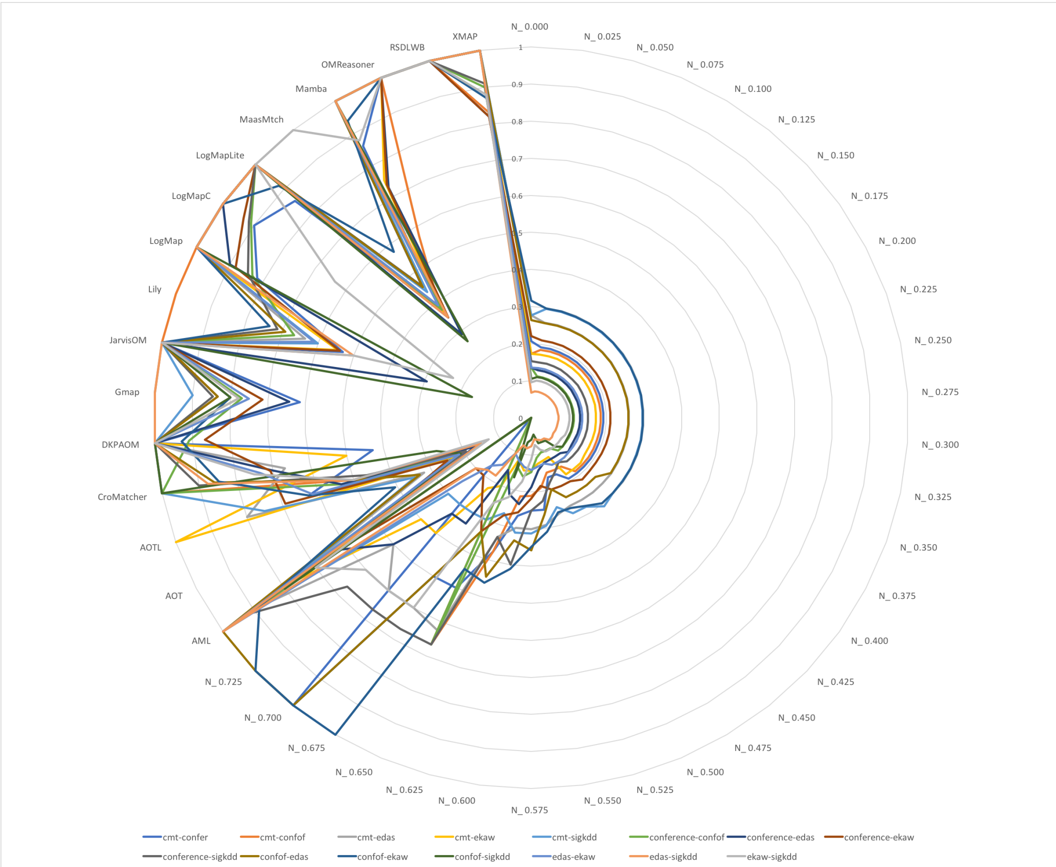


FIGURE A.20: Precision for all ontologies using DbMN_5 compared to the current alignment systems

The alignment results for these current systems, were taken from the OAEI reference alignments for the corresponding dataset pairs. These results show the best performance in comparison to the current systems for the precision of the DbMN_5 approach is between the $\epsilon_n = 0.575$ to 0.25 where the precision values for the ontologies are the highest in comparison to the current systems.

Recall:

The Graph A.21 shows the results of the recall from all the pairs of ontologies, using the DbMN_5 system, over all the incriminations of the N_threshold value.

These recall values are calculated against some of those found by the systems discussed in Chapter 3 including: AML, AOT, AOTL, CROMatcher, DKPAOM, GMap, JarvisOM, Lily, MassMatch, Mamba, OMReasoner, RSDLWB and XMAP.

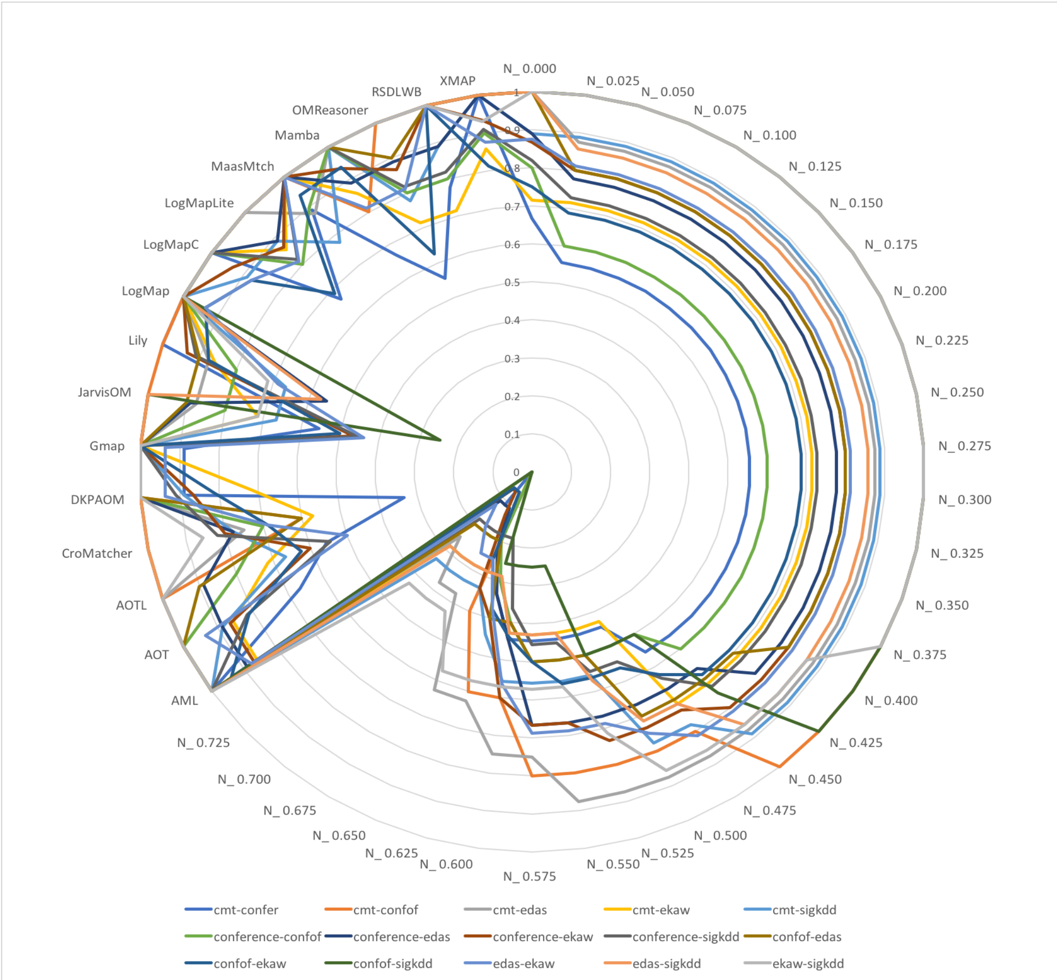


FIGURE A.21: Recall for all ontologies using DbMN_5 compared to the current alignment systems

The alignment results for these current systems, were taken from the OAEI reference alignments for the corresponding dataset pairs. These results show the best performance in comparison to the current systems for the recall of the DbMN_5 system is between the $\epsilon_n = 0.000$ to 0.450 where the recall for the ontologies begins decrease to a value of 0.

A.1.4 Number of concepts and the percentage % of ontology O' shared

One of the research questions evaluating the DbMN approach is the ability for the opponent agent to maintain a level of privacy regarding their assigned ontology (O'). It has been discussed that privacy for the proponent is not possible due to the restriction put on the signature to be mapped in this evaluation of the DbMN approach. This restriction includes all of the proponent's concepts in the ontology O to be included in this signature which needs to be mapped within in the dialogue between the two agents. As this signature includes all of the concepts, there is not possibility for the proponent to not share their full ontology as this is done incrementally in the *initiate* move.

Number of concepts and the percentage % of ontology O' shared for DbMN_5

Table A.1 illustrates the amount of the ontologies that are shared by each of the agents they are assigned to. The table shows the number of concepts shared, and the % of the ontology shared within the ontology assigned to the opponent agent for all the dataset pairs included in the experimentation.

Number of concepts and the percentage % of ontology O' shared for DbMN_6

Table A.2 illustrates both the number of concepts shared, and the percentage of the ontology O' shared. This table shows these results for all the dataset pairs included in the experimentation for the DbMN_6 variant of the approach.

ontologies O-$O\prime$	ϵ_n= 0.000		ϵ_n= 0.575		ϵ_n=0.600		ϵ_n=0.625		ϵ_n=0.650		ϵ_n=0.675		ϵ_n=0.700		ϵ_n=0.725		ϵ_n=0.750		ϵ_n=0.100	
	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%
cmt-**conference**	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	25	41.67	25	41.67	25	41.67
cmt-**confof**	23	58.97	23	58.97	23	58.97	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41
cmt-**edas**	33	31.73	33	31.73	33	31.73	32	30.77	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81
cmt-**ekaw**	30	40.54	27	36.49	27	36.49	25	33.78	27	36.49*	27	36.49	27	36.49	27	36.49	27	36.49	27	36.49
cmt-**sigkdd**	21	42	22	44	22	44	23	46*	23	46	23	46	23	46	23	46	23	46	23	46
conference-**confof**	31	78.48	31	79.49	31	79.49	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92
conference-**edas**	50	48.08	44	42.31	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38
conference-**ekaw**	50	67.57	48	64.86	48	64.86	46	62.16	46	62.16	45	60.81	45	60.81	45	60.81	45	60.81	45	60.81
conference-**sigkdd** 2	32	64	30	60*	30	60	30	60	30	60	30	60	30	60	30	60	30	60	30	60
confof-**edas**	37	35.58	37	35.58	37	35.58	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62
confof-**ekaw**	36	64	35	47.30	33	44.59	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24
confof-**sigkdd**	24	48.65	26	52*	26	52	26	52	26	52	26	52	26	52	26	52	26	52	26	52
edas-**ekaw**	51	68.92	49	66.21	48	64.86	48	64.86	47	63.51	46	62.16	46	62.16	46	62.16	46	62.16	46	62.16
edas-**sigkdd**	39	78	38	76	37	74	36	72	35	70	35	70	35	70	35	70	35	70	35	70
ekaw-**sigkdd**	37	74	37	74	34	68	34	68	33	66	33	66	33	66	33	66	33	66	33	66

Number of classes shared as number of (#) and % of $O\prime$ where ϵ_n =

TABLE A.1: For all ontologies, comprehensive data on number of concepts shared from $O\prime$ with DbMN_5

	Number of classes shared as number of (#) and % of $O\prime$ where ϵ_n =																			
	ϵ_n= 0.000		ϵ_n= 0.575		ϵ_n=0.600		ϵ_n=0.625		ϵ_n=0.650		ϵ_n=0.675		ϵ_n=0.700		ϵ_n=0.725		ϵ_n=0.750		ϵ_n=0.100	
ontologies O-$O\prime$	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%
cmt-**conference**	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	25	41.67	25	41.67	25	41.67
cmt-**confof**	23	58.97	23	58.97	23	58.97	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41
cmt-**edas**	33	31.73	33	31.73	33	31.73	32	30.77	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81
cmt-**ekaw**	30	40.54	27	36.49	27	36.49	25	33.78	27	36.49*	27	36.49	27	36.49	27	36.49	27	36.49	27	36.49
cmt-**sigkdd**	21	42	22	44	22	44	23	46*	23	46	23	46	23	46	23	46	23	46	23	46
conference-**confof**	31	78.48	31	79.49	31	79.49	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92
conference-**edas**	50	48.08	44	42.31	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38
conference-**ekaw**	50	67.57	48	64.86	48	64.86	46	62.16	46	62.16	45	60.81	45	60.81	45	60.81	45	60.81	45	60.81
conference-**sigkdd** 2	32	64	30	60*	30	60	30	60	30	60	30	60	30	60	30	60	30	60	30	60
confof-**edas**	37	35.58	37	35.58	37	35.58	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62
confof-**ekaw**	36	64	35	47.30	33	44.59	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24
confof-**sigkdd**	24	48.65	26	52*	26	52	26	52	26	52	26	52	26	52	26	52	26	52	26	52
edas-**ekaw**	51	68.92	49	66.21	48	64.86	48	64.86	47	63.51	46	62.16	46	62.16	46	62.16	46	62.16	46	62.16
edas-**sigkdd**	39	78	38	76	37	74	36	72	35	70	35	70	35	70	35	70	35	70	35	70
ekaw-**sigkdd**	37	74	37	74	34	68	34	68	33	66	33	66	33	66	33	66	33	66	33	66

TABLE A.2: Exhaustive illustration of shared concepts, across all ontologies for system DbMN_6

A.2 DbMN_7 Experimentation Results

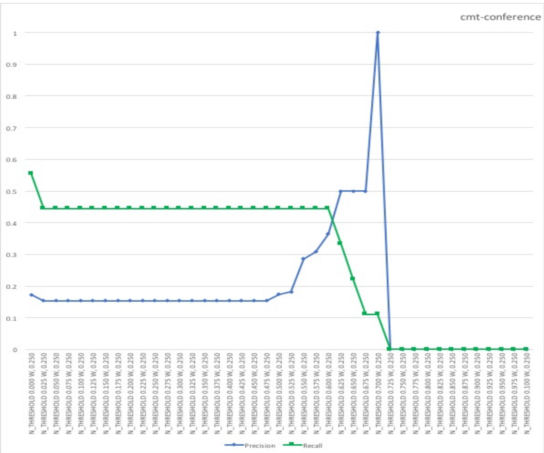
The results as presented in this section, show two graphs for each dataset pairs, as detailed in Chapter 8 for the DbMN_7 variant of the approach. These graphs show the x axis marking the incremented neighbourhood threshold values, and the y axis shows the precision and recall values for the left graph, and the number of mappings on the right.

These results have been discussed in Chapter 8 and are included in this appendix to illustrate each of the results for the dataset pairs individually, to accompany the combined results presented in that chapter.

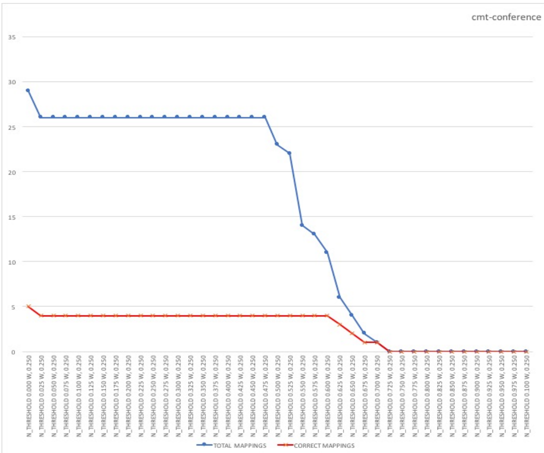
This section will also present the results for the number of concept shared from the opponent agent's assigned ontology (O').

R.22 **For the ontologies O = cmt, O' = conference:**

Graphs illustrating precision and recall values (left) for O = cmt, O' = conference, with the highest precision = 1 and the highest recall = 0.56. Secondly the total mappings found (right) from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, O' = conference, from the values 5 mappings to 0.



(a) O = cmt, O' = conference Precision and Recall

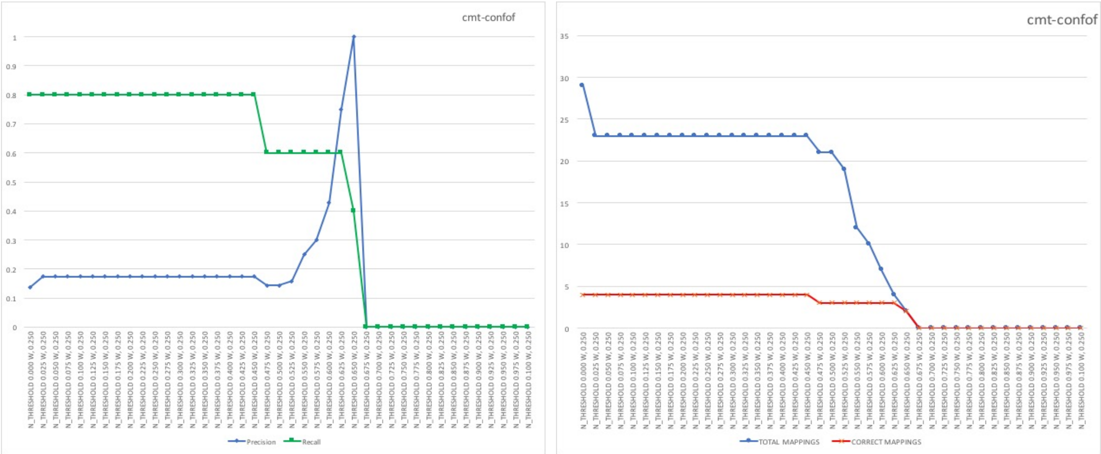


(b) O = cmt, O' = conference Total Mappings

FIGURE A.22: Results for O = cmt, O' = conference

R.23 **For the ontologies O = cmt, O' = confof:**

Graphs illustrating, firstly on the left Precision and Recall values for O = cmt, O' = confof from ϵ_n=[0..1], with the highest precision = 1 and the highest recall = 0.80. Secondly on the right the total mappings found from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, O' = confof, from the values 4 mappings to 0.

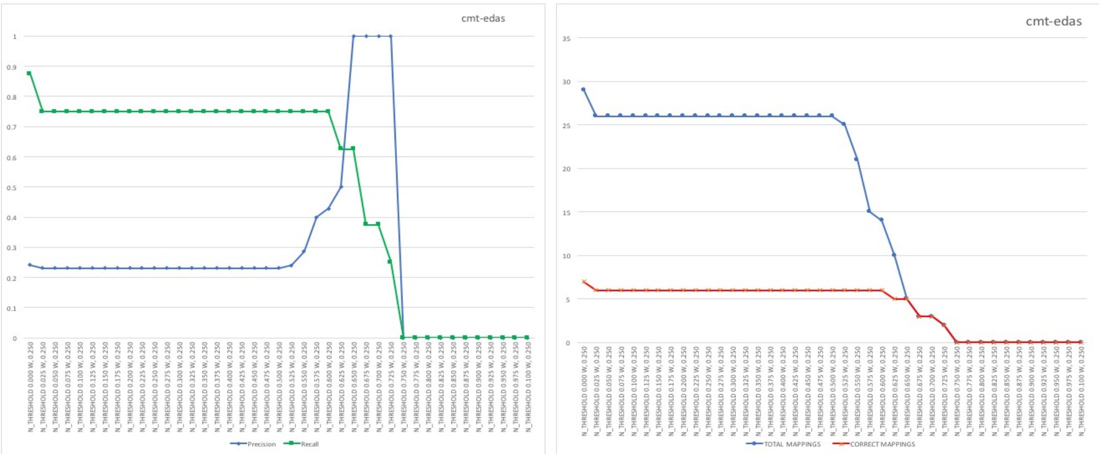


(a) O = cmt, O' = confof Precision and Recall (b) O = cmt, O' = confof Total Mappings

FIGURE A.23: Results for O = cmt, O' = confof

R.24 **For the ontologies O = cmt, O' = edas:**

Graphs illustrating, firstly on the left Precision and Recall values for O = cmt, O' = edas from ϵ_n=[0..1], with the highest precision = 1 and the highest recall = 0.88. Secondly on the right the total mappings found from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, O' = edas, from the values 7 mappings to 0.

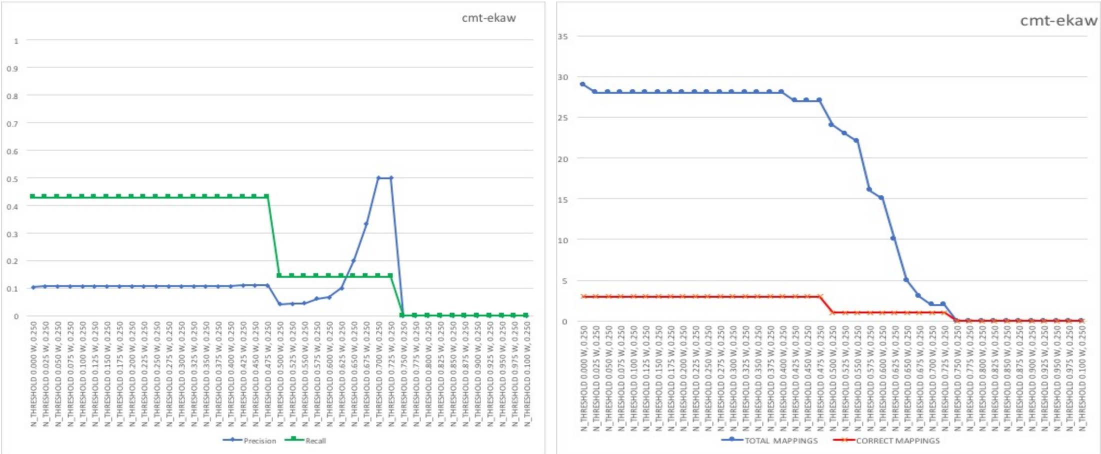


(a) O = cmt, O' = edas Precision and Recall (b) O = cmt, O' = edas Total Mappings

FIGURE A.24: Results for O = cmt, O' = edas

R.25 **For the ontologies O = cmt, O' = ekaw:**

Graphs illustrating, firstly on the left Precision and Recall values for O = cmt, O' = ekaw from ϵ_n=[0..1], with the highest precision = 0.50 and the highest recall = 0.43. Secondly on the right the total mappings found from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, O' = ekaw, from the values 3 mappings to 0.

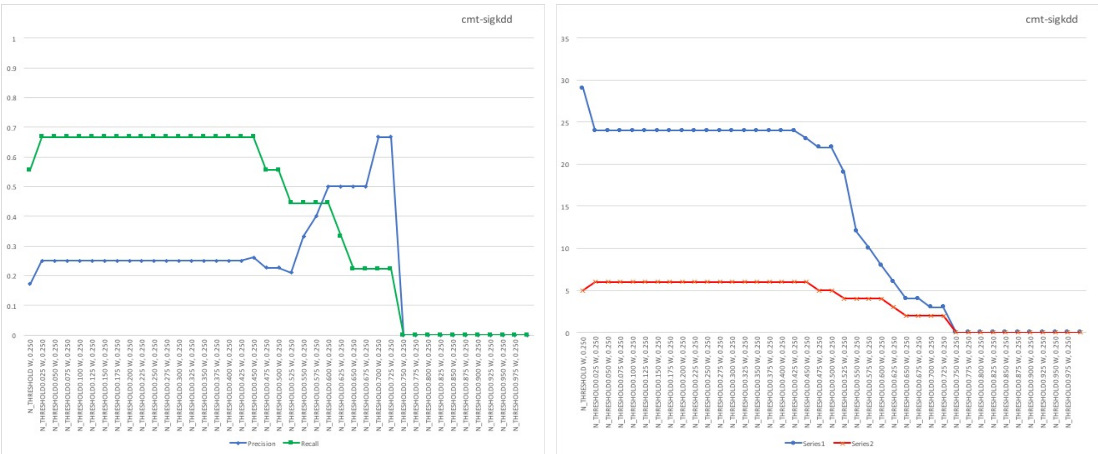


(a) O = cmt, $O\prime$ = ekaw Precision and Recall (b) O = cmt, $O\prime$ = ekaw Total Mappings

FIGURE A.25: Results for O = cmt, $O\prime$ = ekaw

R.26 **For the ontologies O = cmt, $O\prime$ = sigkdd:**

Graphs illustrating, firstly on the left Precision and Recall values for O = cmt, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.667 and the highest recall = 0.56. Secondly on the right the total mappings found from 29 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = cmt, $O\prime$ = sigkdd, from the values 5 mappings to 0.

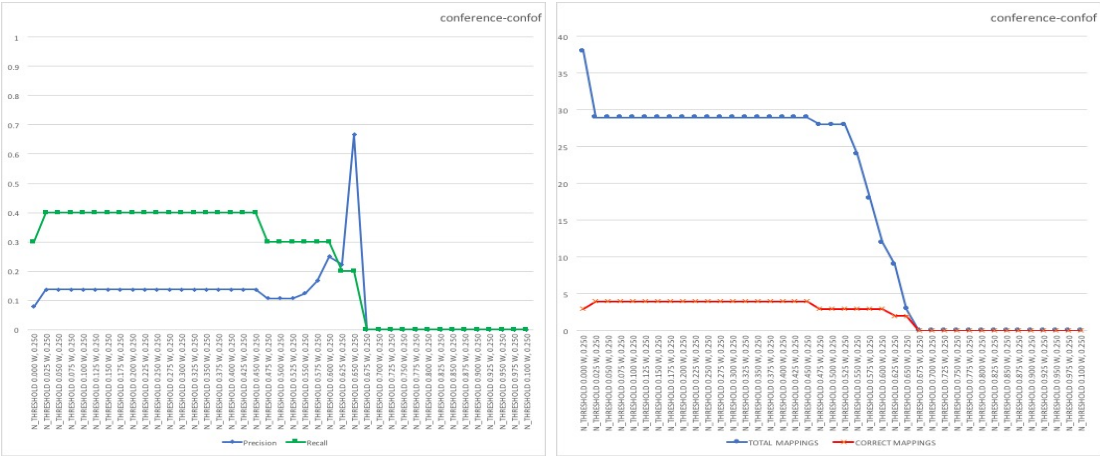


(a) O = cmt, $O\prime$ = sigkdd Precision and Recall (b) O = cmt, $O\prime$ = sigkdd Total Mappings

FIGURE A.26: Results for O = cmt, $O\prime$ = sigkdd

R.27 **For the ontologies O = conference, $O\prime$ = confof:**

Graphs illustrating, firstly on the left Precision and Recall values for O = conference, $O\prime$ = confof from ϵ_n=[0..1], with the highest precision = 0.67 and the highest recall = 0.40. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, $O\prime$ = confof, from the values 4 mappings to 0.



(a) O = conference, $O\prime$ = confof Precision and Recall (b) O = conference, $O\prime$ = confof Total Mappings

FIGURE A.27: Results for O = conference, $O\prime$ = confof

R.28 **For the ontologies O = conference, $O\prime$ = edas:**

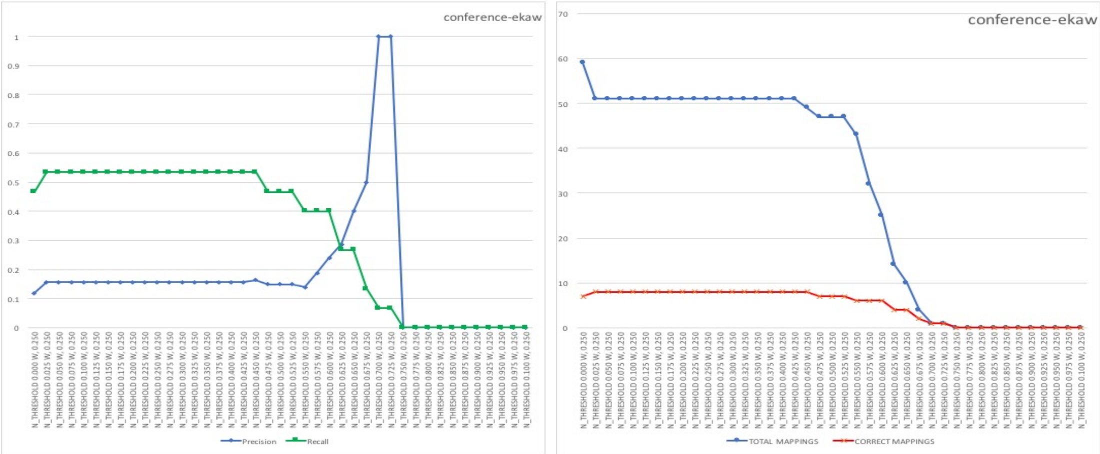
Graphs illustrating, firstly on the left Precision and Recall values for O = conference, $O\prime$ = edas from ϵ_n=[0..1], with the highest precision = 0.25 and the highest recall = 0.56. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, $O\prime$ = edas, from the values 5 mappings to 0.

(a) O = conference, $O\prime$ = edas Precision and Recall (b) O = conference, $O\prime$ = edas Total Mappings

FIGURE A.28: Results for O = conference, $O\prime$ = edas

R.29 **For the ontologies O = conference, $O\prime$ = ekaw:**

Graphs illustrating, firstly on the left Precision and Recall values for O = conference, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 1 and the highest recall = 0.47. Secondly on the right the total mappings found from 59 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, O' = ekaw, from the values 7 mappings to 0.



(a) O = conference, O' = ekaw Precision and Recall (b) O = conference, O' = ekaw Total Mappings

FIGURE A.29: Results for O = conference, O' = ekaw

R.30 **For the ontologies O = conference, O' = sigkdd:**

Graphs illustrating, firstly on the left Precision and Recall values for O = conference, O' = sigkdd from ϵ_n=[0..1], with the highest precision = 0.50 and the highest recall = 0.36. Secondly on the right the total mappings found from 49 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = conference, O' = sigkdd, from the values 4 mappings to 0.

(a) O = conference, O' = sigkdd Precision and Recall (b) O = conference, O' = sigkdd Total Mappings

FIGURE A.30: Results for O = conference, O' = sigkdd

R.31 **For the ontologies O = confof, O' = edas:**

Graphs illustrating, firstly on the left Precision and Recall values for O = confof, O' = edas from ϵ_n=[0..1], with the highest precision = 1and the highest recall

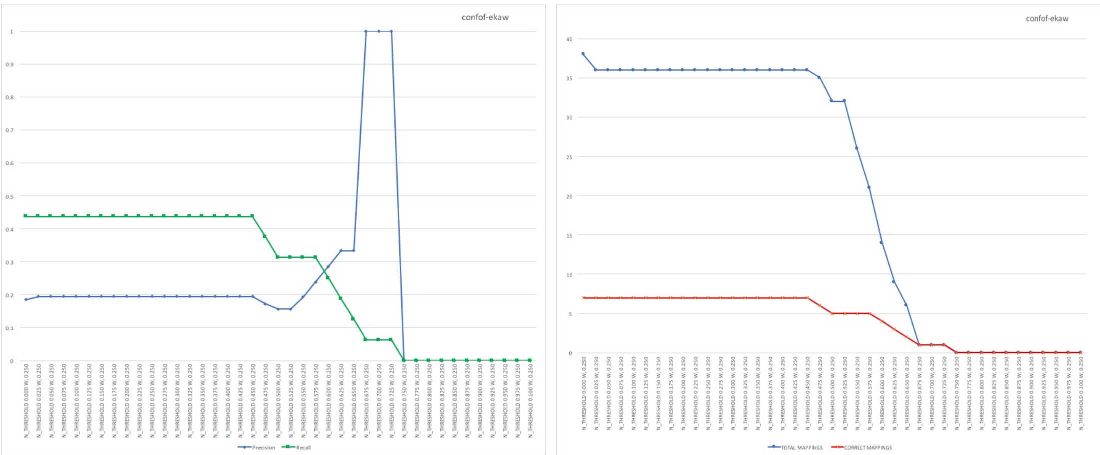
= 0.60. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = confof, $O\prime$ = edas, from the values 6 mappings to 0.

(a) O = confof, $O\prime$ = edas Precision and Recall (b) O = confof, $O\prime$ = edas Total Mappings

FIGURE A.31: Results for O = confof, $O\prime$ = edas

R.32 **For the ontologies O = confof, $O\prime$ = ekaw:**

Graphs illustrating, firstly on the left Precision and Recall values for O = confof, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 1 and the highest recall = 0.44. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = confof, $O\prime$ = ekaw, from the values 7 mappings to 0.



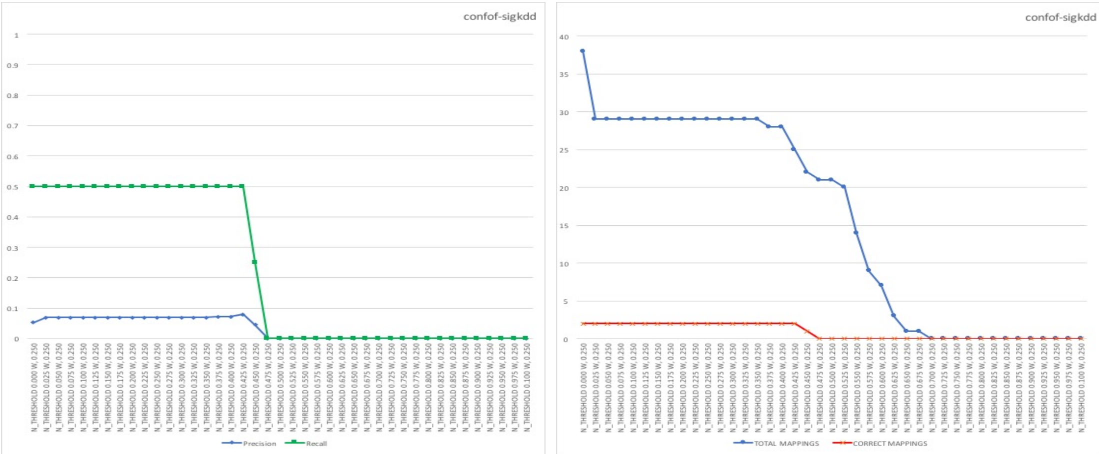
(a) O = confof, $O\prime$ = ekaw Precision and Recall (b) O = confof, $O\prime$ = ekaw Total Mappings

FIGURE A.32: Results for O = confof, $O\prime$ = ekaw

R.33 **For the ontologies O = confof, $O\prime$ = sigkdd:**

Graphs illustrating, firstly on the left Precision and Recall values for O = confof, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.08 and the highest recall

= 0.50. Secondly on the right the total mappings found from 38 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = confof, $O\prime$ = sigkdd, from the values 2 mappings to 0.

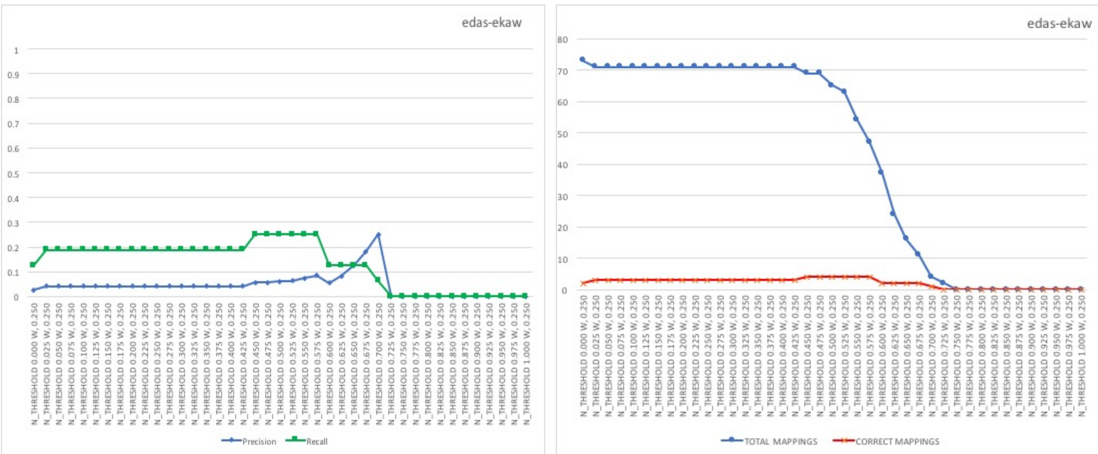


(a) O = confof, $O\prime$ = sigkdd Precision and Recall (b) O = confof, $O\prime$ = sigkdd Total Mappings

FIGURE A.33: Results for O = confof, $O\prime$ = sigkdd

R.34 **For the ontologies O = edas, $O\prime$ = ekaw:**

Graphs illustrating, firstly on the left Precision and Recall values for O = edas, $O\prime$ = ekaw from ϵ_n=[0..1], with the highest precision = 0.25 and the highest recall = 0.19. Secondly on the right the total mappings found from 73 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = edas, $O\prime$ = ekaw, from the values 4 mappings to 0.



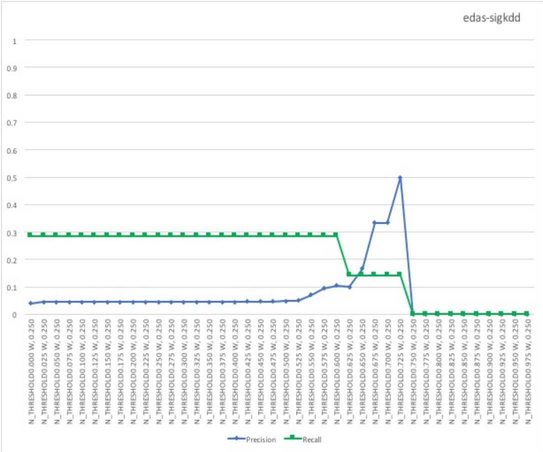
(a) O = edas, $O\prime$ = ekaw Precision and Recall (b) O = edas, $O\prime$ = ekaw Total Mappings

FIGURE A.34: Results for O = edas, $O\prime$ = ekaw

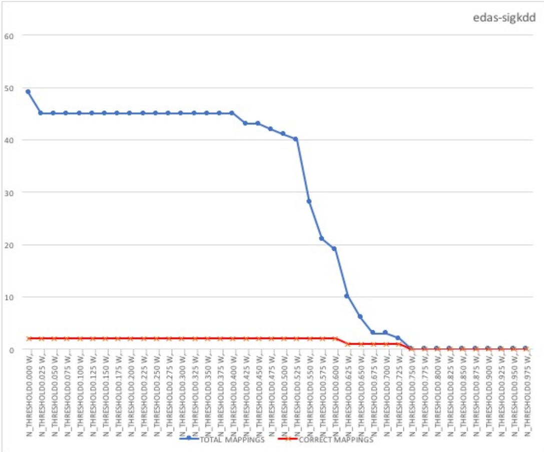
R.35 **For the ontologies O = edas, $O\prime$ = sigkdd:**

Graphs illustrating, firstly on the left Precision and Recall values for O = edas, $O\prime$ = sigkdd from ϵ_n=[0..1], with the highest precision = 0.50 and the highest recall

= 0.29. Secondly on the right the total mappings found from 49 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = edas, O' = sigkdd, from the values 2 mappings to 0.



(a) O = edas, O' = sigkdd Precision and Recall

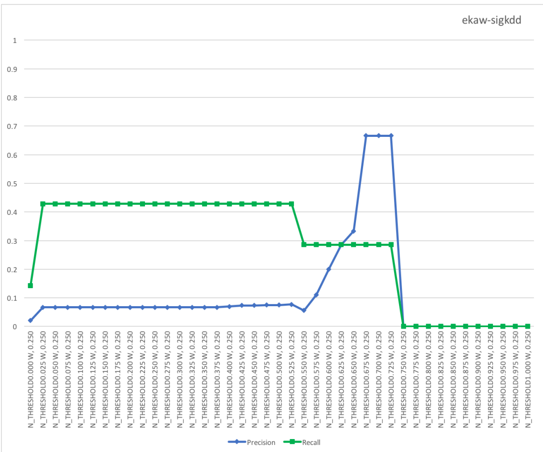


(b) O = edas, O' = sigkdd Total Mappings

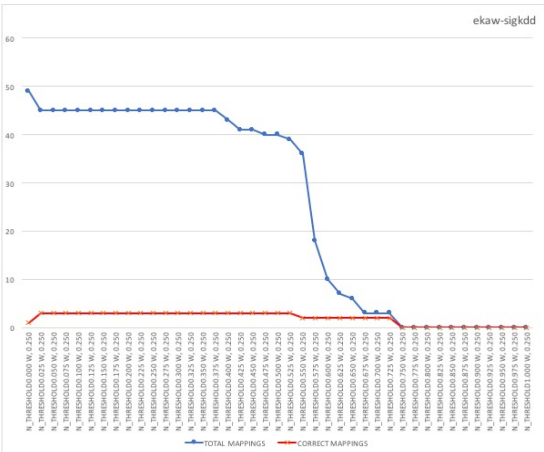
FIGURE A.35: Results for O = edas, O' = sigkdd

R.36 **For the ontologies O = ekaw, O' = sigkdd:**

Graphs illustrating, firstly on the left Precision and Recall values for O = ekaw, O' = sigkdd from ϵ_n=[0..1], with the highest precision = 0.667 and the highest recall = 0.143. Secondly on the right the total mappings found from 49 to 0, and the mappings found by the system that feature in the *Platinum Standard* i.e. 'correct mappings' between O = ekaw, O' = sigkdd, from the values 3 mappings to 0.



(a) O = ekaw, O' = sigkdd Precision and Recall



(b) O = ekaw, O' = sigkdd Total Mappings

FIGURE A.36: Results for O = ekaw, O' = sigkdd

Number of concepts and the percentage % of ontology O' shared for DbMN_7

Table A.3 illustrates the amount of the ontologies that are shared by each of the agents they are assigned to. The table shows the number of concepts shared, and the percentate % of the ontology shared within the ontology assigned to the opponent agent for all the dataset pairs included in the evaluation.

	Number of classes shared as number of (#) and % of $O\prime$ where ϵ_n =																			
	ϵ_n= 0.000		ϵ_n= 0.575		ϵ_n=0.600		ϵ_n=0.625		ϵ_n=0.650		ϵ_n=0.675		ϵ_n=0.700		ϵ_n=0.725		ϵ_n=0.750		ϵ_n=0.100	
ontologies O-$O\prime$	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%	# c	%
cmt-**conference**	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	26	43.33	25	41.67	25	41.67	25	41.67
cmt-**confof**	23	58.97	23	58.97	23	58.97	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41	22	56.41
cmt-**edas**	33	31.73	33	31.73	33	31.73	32	30.77	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81	31	29.81
cmt-**ekaw**	30	40.54	27	36.49	27	36.49	25	33.78	27	36.49	27	36.49	27	36.49	27	36.49	27	36.49	27	36.49
cmt-**sigkdd**	21	42	22	44	22	44	23	46	23	46	23	46	23	46	23	46	23	46	23	46
conference-**confof**	31	78.48	31	79.49	31	79.49	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92	30	76.92
conference-**edas**	50	48.08	44	42.31	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38	42	40.38
conference-**ekaw**	50	67.57	48	64.86	48	64.86	46	62.16	46	62.16	45	60.81	45	60.81	45	60.81	45	60.81	45	60.81
conference-**sigkdd**	32	64	30	60	30	60	30	60	30	60	30	60	30	60	30	60	30	60	30	60
confof-**edas**	37	35.58	37	35.58	37	35.58	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62	36	34.62
confof-**ekaw**	36	64	35	47.30	33	44.59	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24	32	43.24
confof-**sigkdd**	24	48.65	26	52*	26	52	26	52	26	52	26	52	26	52	26	52	26	52	26	52
edas-**ekaw**	51	68.92	49	66.21	48	64.86	48	64.86	47	63.51	46	62.16	46	62.16	46	62.16	46	62.16	46	62.16
edas-**sigkdd**	39	78	38	76	37	74	36	72	35	70	35	70	35	70	35	70	35	70	35	70
ekaw-**sigkdd**	37	74	37	74	34	68	34	68	33	66	33	66	33	66	33	66	33	66	33	66

Table A.3: For all ontologies, comprehensive data on number of concepts shared from $O\prime$ with DbMN_7

A.3 DbMN_7 Experimentation Results

As outlined in Chapter 8 the f-measure was calculated as an evaluation metric for each of the dataset pairs used in the experimentation. Figure A.37 illustrates an average f-measure for each ϵ_n bound from [0..1] over each of the dataset results. This provides an illustration of how the f-measure varies over the threshold bound for the three variants.

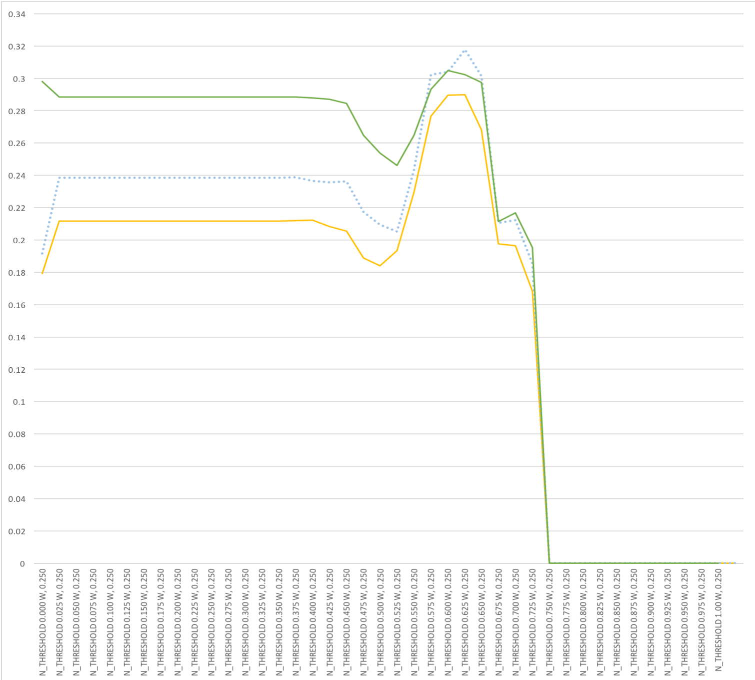


FIGURE A.37: Averaged F-measure values across all the ontologies, for ordered and unordered signature of DbMN_7 and DbMN_6 and DbMN_5 (presented as a single series.)

In Figure A.37 the average f-measure for the DbMN_5 and DbMN_6 are presented with the green line. The DbMN_7 f-measure values have been separated into the results where the signature Σ^t was an random list (blue dotted line), and an ordered alphabetic list as the yellow line. The averaged f-measure for all the variants of the dialogue approach illustrate that the optimum f-measure is at the $\epsilon_n = 0.575$ and 0.650. It is at $\epsilon_n = 0.750$ for all variants that the f-measure reduces to a value of 0, in corroboration with the previously detailed recall and precision values.

Appendix B

Dataset Benchmarks

Chapter Outline

This appendix (B) documents all of the independent benchmarks, for the 15 experiment pairs of ontologies run throughout the implementation of the systems presented in this thesis. These benchmarks state the original gold standard benchmarks, and the pruned platinum standard benchmarks, used into order to empirically evaluate the approach presented in Chapter 8.

B.1 DbMN System Benchmarks

As detailed in Chapter 8 the approach has been evaluated using precision and recall values comparing the alignments generated using the various dataset pairs, and compared to a benchmark alignment. In the OAEI this alignment is a gold standard alignment consisting of classes and properties. As discussed in Chapter 8 the evaluation of the DbMN approach uses a pruned 'platinum standard' as a reference alignment to evaluate the generated alignments produced by the variants of the approach.

The number of entities in the both the OAEI *gold standard* benchmark taken from the reference alignment [67], and the *platinum standard* generated in the implementation for all the variants of the approach DbMN_5, DbMN_6 and DbMN_7 can be seen in Table B.1.

Table B.1: Number of entities in each of benchmark alignments across all the dataset pairs, for the variants 5, 6 and 7.

Datasets		Benchmarks	
Ontology O	Ontology O'	Gold Standard	Platinum Standard
cmt	conference	15	9
cmt	confof	16	5
cmt	edas	13	8
cmt	ekaw	11	7
cmt	sigkdd	12	9
conference	confof	15	10
conference	edas	17	9
conference	ekaw	27	15
conference	sigkdd	15	11
confof	edas	22	10
confof	ekaw	20	16
confof	sigkdd	7	4
edas	ekaw	29	16
edas	sigkdd	15	7
ekaw	sigkdd	11	7

As detailed in Chapter 8, the variants of the approach implemented, have been evaluated in comparison to a benchmark standard. This benchmark has been pruned from the 'gold standard' benchmark, used in the OAEI, and has removed all mappings containing property concepts, and has only included mappings that were found by the majority of the current systems.

These current systems some of which have been discussed in Chapter 3 and include: AML, AOT, AOTL, CROMatcher, DKPAOM, GMap, JarvisOM, Lily, MassMatch, Mamba, OMReasoner, RSDLWB and XMAP. Mappings were added into the pruned platinum standard, if they were consistently found by the majority of these systems, providing a potentially more accurate alignment in which to use as an evaluation metric for the systems implemented, taking into account that the DbMN systems only match class concepts. Across all the variants the benchmark for the alignments is the same, and as detailed in the experiment preliminaries of Chapter 8. This section will detail each of the dataset pairs used in the experimentation, to state which concepts are included in the platinum standard for this implementation, and also which have been added or removed from the gold standard. The Tables B.2, B.3 and B.4 can be read to include the following details:

- (−) if the mapping is removed from the gold standard, and thus not included in the platinum standard;
- (+) if the mapping is included in the platinum standard and does not appear in the gold standard;
- and no marking, if the mapping is included in both the platinum and gold standard benchmarks.

TABLE B.2: Benchmarks for the individual dataset pairs, for all systems

O= cmt $O\prime$= conference

Gold Standard	Platinum Standard
Conference → Conference_volume (−)	Conference → Conference (+)
Preference → Review_preference (−)	ProgramCommitteeMember → Committee_member (+)
Author → Regular_author (−)	Review → Review
Person → Person	Reviewer → Reviewer (+)
email → has_an_email (−)	Co-author → Contribution_co-author
Co-author → Contribution_co-author	Paper → Paper
PaperAbstract → Abstract (−)	Person → Person
Document → Conference_document	Document → Conference_Document
Review → Review	ProgramCommittee → Program_Committee
Conference → Conference	
ProgramCommittee → Program_committee	
Chairman → Chair (−)	
SubjectArea → Topic (−)	
assignedByReviewer → invited_by (−)	
assignExternalReviewer → invites_co-reviewers (−)	

O= cmt $O\prime$= confof

Gold Standard	Platinum Standard
ProgramCommitteeChair → Chair_PC (−)	Person → Person
writePaper → writes (−)	Author → Author
Author → Author	Administrator → Administrator
ConferenceMember → Member (−)	Paper → Paper (+)
Administrator → Administrator	Conference → Conference
title → hasTitle (−)	
SubjectArea → Topic (−)	
PaperFullVersion → Paper (−)	
hasBeenAssigned → reviewes (−)	
hasAuthor → writtenBy (−)	
Conference → Conference	
ProgramCommitteeMember → Member_PC (−)	
hasSubjectArea → dealsWith (−)	
Person → Person	
Paper → Contribution (−)	
email → hasEmail (−)	

O= cmt $O\prime$= edas

Gold Standard	Platinum Standard
Person → Person	Author → Author
Conference → Conference	ConferenceChair → ConferenceChair
Author → Author	Paper → Paper
hasConferenceMember → hasMember (−)	Document → Document
memberOfConference → isMemberOf (−)	Conference → Conference
ConferenceChair → ConferenceChair	Reviewer → Reviewer (+)
Review → Review	Person → Person
Document → Document	Review → Review
Paper → Paper	
hasAuthor → isWrittenBy (−)	
hasBeenAssigned → isReviewing (−)	
assignedTo → isReviewedBy (−)	

O= cmt $O\prime$= ekaw

Gold Standard	Platinum Standard
Document → Document	Author → Paper_Author
ConferenceMember → Conference_Participant(−)	Document → Document
Author → Paper_Author	Review → Review
writtenBy → reviewWrittenBy (−)	Person → Person
hasBeenAssigned → reviewerOfPaper (−)	Paper → Paper
Person → Person	Conference → Conference
Conference → Conference	Reviewer → Possible_Reviewer (+)
assignedTo → hasReviewer (−)	
Review → Review	
Paper →Paper	
PaperFullVersion → Regular_Paper (−)	

O= cmt $O\prime$= sigkdd

Gold Standard	Platinum Standard
Conference → Conference	Person → Person
Paper → Paper	ProgramCommittee → Program_Committee
ProgramCommitteeMember → Program_Committee_member	ProgramCommitteeMember → Program_Committee_member
Document → Document	Conference → Conference
ConferenceChair → General_Chair (−)	Paper → Paper
email → E-mail (−)	Document → Document
Review → Review	Review → Review
ProgramCommittee → Program_Committee	ProgramCommitteeChair → Program_Chair
ProgramCommitteeChair → Program_Chair	Author → Author
Author → Author	
submitPaper → submit (−)	
Person → Person	

O= conference $O\prime$= confof

Gold Standard	Platinum Standard
Conference_participant → Participant	Conference_participant → Participant
has_an_email → hasEmail (−)	Conference_contribution → Contribution
Poster → Poster	Workshop → Workshop
Organization → Organization	Person → Person
Topic → Topic	Tutorial → Tutorial
Workshop → Workshop	Topic → Topic
Paper → Paper	Poster → Poster
Person → Person	Conference → Conference
Conference_contribution → Contribution	Paper → Paper
Tutorial → Tutorial	Organization → Organization
Conference_volume → Conference (−)	
has_a_track-workshop-tutorial_topic → hasTopic (−)	
Regular_author → Author (−)	
has_the_last_name → hasSurname (−)	
has_the_first_name → hasFirstName (−)	

Table B.3: Benchmarks for the individual dataset pairs, for all systems

O= conference O'= edas

Gold Standard	*Platinum Standard*
Person → Person	Topic → Topic
Conference_participant → Attendee (−)	Workshop → Workshop
Organization → Organization	Organization → Organization
Reviewer → Reviewer	Reviewer → Reviewer
has_the_first_name → hasFirstName (−)	Paper → Paper
Conference_part → ConferenceEvent (−)	Person → Person
Workshop → Workshop	Review → Review (+)
Conference_document → Document	Conference → Conference
Paper → Paper	Conference_document → Document
has_a_review_expertise → hasRating (−)	
has_the_last_name → hasLastName (−)	
Review → Review	
Conference_volume → Conference (−)	
Rejected_contribution → RejectedPaper (−)	
Topic → Topic	
Accepted_contribution → AcceptedPaper (−)	
Regular_author → Author (−)	

O= conference O'= ekaw

Gold Standard	*Platinum Standard*
Information_for_participants → Programme_Brochure (−)	Invited_talk → Invited_Talk (+)
Person → Person	Abstract → Abstract
Tutorial → Tutorial	Review → Review
Review → Review	Conference → Conference (+)
has_a_review → hasReview (−)	Track → Track
Workshop → Workshop	Paper → Paper (+)
Late_paid_applicant → Late-Registered_Participant (−)	Invited_speaker → Invited_Speaker
Early_paid_applicant → Early-Registered_Participant (−)	Reviewer → Possible_Reviewer (+)
Organization → Organisation	Person → Person
Track-workshop_chair → Workshop_Chair (−)	Organization → Organisation
Abstract → Abstract	Conference_document → Document
Conference_proceedings → Conference_Proceedings	Workshop → Workshop
Conference_volume → Conference (−)	Conference_participant → Conference_Participant (+)
Rejected_contribution → Rejected_Paper (−)	Tutorial → Tutorial
Poster → Poster_Paper (−)	Conference_proceedings → Conference_Proceedings
Track → Track	
Topic → Research_Topic (−)	
Conference_www → Web_Site (−)	
Invited_speaker → Invited_Speaker	
contributes → authorOf (−)	
Accepted_contribution → Accepted_Paper (−)	
Conference_document → Document (−)	
Reviewed_contribution → Evaluated_Paper (−)	
Submitted_contribution → Submitted_Paper (−)	
Regular_author → Paper_Author (−)	

O= conference O'= sigkdd

Gold Standard	*Platinum Standard*
Abstract → Abstract (−)	Person → Person (+)
Invited_speaker → Invited_Speaker	Organizing_committee → Organizing_Committee
Regular_author → Author (−)	Abstract → Abstract
Review → Review	Invited_speaker → Invited_Speaker
Program_committee → Program_Committee	Review → Review
Conference_volume → Conference (−)	Committee → Committee
Conference_fees → Fee (−)	Conference → Conference (+)
has_an_email → E-mail (−)	Paper → Paper
Paper → Paper	Conference_document → Document
Organizing_committee → Organizing_Committee	Conference_fees → Fee
Person → Person	Program_committee → Program_Committee
Committee → Committee	
is_given_by → presentationed_by (−)	
gives_presentations → presentation (−)	
Conference_document → Document	

O= confof O'= edas

Gold Standard	*Platinum Standard*
Trip → Excursion (−)	Author → Author
Social_event → SocialEvent	Organization → Organization
reviewes → isReviewing (−)	Person → Person
Organization → Organization	Country → Country
writtenBy → isWrittenBy (−)	Topic → Topic
Working_event → AcademicEvent (−)	Social_event → SocialEvent
Reception → Reception	Conference → Conference (+)
hasSurname → hasLastName (−)	Reception → Reception
Workshop → Workshop	Paper → Paper
Author → Author	Workshop → Workshop
hasFirstName → hasFirstName (−)	
Event → ConferenceEvent (−)	
Topic → Topic	
Country → Country	
Participant → Attendee (−)	
Person → Person	
Member_PC → TPCMember (−)	
Paper → Paper	
writes → hasRelatedPaper (−)	

Table B.4: Benchmarks for the individual dataset pairs, for all systems

O= **confof** O'= **ekaw**		O= **edas** O'= **ekaw**	
Gold Standard	*Platinum Standard*	*Gold Standard*	*Platinum Standard*
Tutorial → Tutorial	Student → Student	ConferenceDinner → Conference_Banquet (−)	Workshop → Workshop
Poster → Poster_Paper (−)	Banquet → Conference_Banquet	AcademicEvent → Scientific_Event (−)	SessionChair → Session_Chair
Social_event → Social_Event	Conference → Conference	AcceptedPaper → Accepted_Paper	Review → Review
Person → Person	Author → Paper_Author	isReviewedBy → hasReviewer (−)	SocialEvent → Social_Event
Working_event → Scientific_Event (−)	Event → Event	Place → Location (−)	ConferenceSession → Conference_Session (+)
Conference → Conference	Member_PC → PC_Member	AcademiaOrganization → Academic_Institution (−)	Document → Document
Author → Paper_Author	Trip → Trip (+)	SocialEvent → Social_Event	Programme → Programme_Brochure
Banquet → Conference_Banquet	Tutorial → Tutorial	isReviewing → reviewerOfPaper (−)	Paper → Paper
Workshop → Workshop	Organization → Organisation	Organization → Organisation	Author → Paper_Author
Topic → Research_Topic (−)	Chair_PC → PC_Chair	Author → Paper_Author	Organization → Organisation
Contribution → Paper (−)	Workshop → Workshop	isLocationOf → locationOf (−)	Presenter → Presenter (+)
Participant → Conference_Participant	Person → Person	Topic → Research_Topic (−)	Person → Person
Chair_PC → PC_Chair	Participant → Conference_Participant	Document → Document	RejectedPaper → Rejected_Paper
Organization → Organisation	Paper → Paper (+)	RejectedPaper → Rejected_Paper	Reviewer → Possible_Reviewer (+)
Student → Student	University → University	ConferenceEvent → Event (−)	AcceptedPaper → Accepted_Paper
University → University	Social_event → Social_Event	SessionChair → Session_Chair	Conference → Conference (+)
Trip → Conference_Trip (−)		Person → Person	
Member_PC → PC_Member		Programme → Programme_Brochure	
Scholar → Student (−)		Review → Review	
Event → Event		Workshop → Workshop	
		Paper → Paper	
		Attendee → Conference_Participant (−)	
		hasLocation → heldIn (−)	

O= **confof** O'= **sigkdd**		O= **ekaw** O'= **sigkdd**	
Gold Standard	*Platinum Standard*	*Gold Standard*	*Platinum Standard*
Person → Person	Person → Person	Conference → Conference	Document → Document
Member_PC → Program_Committee_member (−)	Author → Author	Person → Person	Invited_Speaker → Invited_Speaker
hasEmail → E-mail (−)	Paper → Paper	Paper → Paper	Person → Person
Author → Author	Conference → Conference	Review → Review	Abstract → Abstract
Conference → Conference		Invited_Speaker → Invited_Speaker (−)	Paper → Paper
Chair_PC → Program_Chair (−)		OC_Member → Organizing_Committee_member (−)	Conference → Conference
Paper → Paper		Abstract → Abstract	Review → Review
		PC_Chair → Program_Chair (−)	
		Paper_Author → Abstract (−)	
		Abstract → Author (−)	
		Document → Document	
		Location → Place (−)	

O= **edas** O'= **sigkdd**			
Gold Standard	*Platinum Standard*		
Place → Place	Document → Document		
hasCostAmount → Price (−)	Author → Author		
Person → Person	Review → Review		
hasName → Name_of_conference (−)	Person → Person		
ConferenceVenuePlace → Conference_hall (−)	Paper → Paper		
Author → Author	Place → Place		
AccommodationPlace → Hotel (−)	Conference → Conference		
startDate → Start_of_conference (−)			
ConferenceChair → General_Chair (−)			
Conference → Conference			
endDate → End_of_conference (−)			
Review → Review			
Document → Document			
Paper → Paper			
Attendee → Listener (−)			

Appendix C

Dialogue Protocol, alternative representation

Chapter Outline

This Appendix (C) documents alternative representation of the Dialogue Protocol, borrowed from the documentation of the PARMA protocol presented in [8, 9].

There are multiple methods in which to represent a dialogical protocol, this section will define the choice of representation of the protocol presented in this thesis, and why this representation has been selected over potential alternatives.

One representation of a protocol is based in the form of temporal logic as seen in [90, 106, 108] which represents the clauses needed by the agent in order to utilise the dialogue defined in both EI, LCC and ConDec protocols respectively.

An alternative representation is in the form of process calculus for example π calculus. As with the temporal logic, π calculus is a very expressive form of representation for presenting a protocol. In the case of the representation of the protocol presented in this thesis, it was decided that these two approaches were unnecessary due to the restrictions and assumptions made within the dialogue presented in Chapter 5. As the DbMN approach aims at verifying that an alignment can be found with no prior knowledge of the agent ontologies, there were a number of simplifying assumptions made to keep the protocol simple. Hence a semi formal presentation was chosen for the protocol.

As presented in Chapter 4 a key element of the protocol is the commitment store shared by the agents, and independently kept by each agent. These two stores are described in detail later in this chapter, and this table indicates when these stores are updated, and cleared through the use of the protocol presented here. It is assumed that once a candidate mapping is accepted and stored in the commitment store, it is not retractable, in such that, the agents can no longer negotiate over an accepted mapping, or attack this once it has been accepted. This restriction has been used in order to assure completeness and termination of the protocol.

The protocol is defined over a finite set of states and legal locutions available to the agents at any state in the dialogue. These available moves and next available moves are detailed in terms of formal pre and post conditions in this chapter, this table presented in this section provides an intuitive axiomatic semantic representation of the protocol presented in this thesis. This table presents the locutions of the DbMN protocol which are shown in the left-most column of table, with the phases of the dialogue in which these locutions appear in bold.

Tables C.1, C.2 and C.3 presents the pre-conditions needed for to allow for each move in the protocol and any post-conditions arising from the use of these available moves. This table has been developed in line with the axiomatic semantic representation of the PARMA Protocol in [8, 9], documenting pre and post conditions for a dialogue game.

The Tables C.1, C.2 and C.3 are divided in to the following phases as detailed in the previous chapter, *Open* and *Propose*, *Confirm* and finally the *Close* phase of the dialogue.

Table C.1 presented here, shows the *Open* and *Propose* phases of the dialogue In these first two phases of the dialogue, the agents are sharing and receiving single concept labels, which are presented as the source C and target $C\prime$ of a potential correspondence M to be found by the agents within this dialogue. The designation of the agent roles as *proponent* and *opponent* is selected at random and the signature to be mapped Σ^t,

Table C.1: Open and Propose phases represented as axiomatic semantics

Locution	*Pre-conditions*	*Post-conditions*
Open		
Enter Dialogue	•No agents currently in the dialogue	•Both agents enter the dialogue and commit to dialogue rules
Initiate (C)	•Proponent has the signature Σ^t, which is not an empty set.	•Proponent has entered the dialogue, and begun negotiation on a concept C which is previously unshared. •The agents are now committed to finding or rejecting a correspondence for this concept.
Propose		
Propose ($C\prime$)	•Opponent calculates the lexically similarity of concept C to its own ontology.	•Opponent shares the most lexically similar concept to C in from its ontology $C\prime$ which is previously unshared.

consists of the concepts within its ontology the proponent agent wishes to find mappings for.

Table C.2: Confirm phase represented as axiomatic semantics

Locution	*Pre-conditions*	*Post-conditions*
Confirm		
Justify (M)	•The receiving agent has received a proposed candidate mapping M, but requires a Premise with enough support $Pr \geq \sigma_n$ to make an assert. •The receiving agent calculates σ_n.	•If the $Pr \geq \sigma_n$ an *assert* can be made. •If $Pr < \sigma_n$ a *testify* is used to gather support.
Testify (M)	•The receiving agent calculates the triple as part of a premise Pr to support. the candidate mapping M	•The sender agent, shares this triple, which has not been previously shared, until there are no further triples to share. •This shared triple when received in this testify move will been committed to the receiving agents Gamma Store Γ as part of a Pr.
Reject	•The receiving agent calculates if the candidate mapping M has enough support to be accepted. •The Proponent agent cannot exceed the threshold level for σ_l if the previous moves a *propose*. •The Proponent agent cannot exceed the σ_n if the previous move is a *testify* or *reject*. •The Opponent agent cannot exceed the σ_n if the previous move is a *testify*.	•The candidate mapping M cannot be accepted due to either σ_l or σ_n not being over threshold.
Assert (M)	•The receiving agent has enough support σ_n in the Pr to assert the candidate mapping M $Pr \geq \sigma_n$.	•The sending agent has enough support to assert the mapping M with a corresponding premise Pr, $Pr \geq \sigma_n$.

Table C.2 presented here, shows the *Confirm* phase of the dialogue. This confirm phase works primarily with $\langle subject, predicate, object \rangle$ triples, where the subject is rooted to the C or $C\prime$ concept respectively depending on the sender agent. These triples are shared to form a premise Pr, in which to provide a structural support for the candidate mapping M.

TABLE C.3: Close phase represented as axiomatic semantics

Locution	*Pre-conditions*	*Post-conditions*
Close		
Accept (M)	•An assert is made on a correspondence M with a supporting premise Pr. •Receiving agent calculates if M can be accepted if the Pr is over or equal to the assigned threshold σ_n value.	•Asserted M is accepted with a Pr. •Dialogue over concept initiated in M is complete, and this dialogue is over. •Next concept is initiated from Σ^t iff $\Sigma^t \neq \emptyset$ •The candidate mapping is committed to and stored in the Commitment Store CS.
End	•Dialogue ended as Σ^t is $\emptyset$	•Both agents exit the dialogue

Table C.2 presented here, shows the *Close* phase of the dialogue and presents the acceptance of a mapping and the ending of the dialogue between the two participating agents. This close phase, can also include the *fail* move seen in the open phase in Table C.2, and presents the failing of a correspondent mapping, and has not been included in this table for the sake of brevity. This close phase, presents the agents accepting a mapping into the alignment generated between them, and sees the mapping M committed to the commitment store CS. As will be formalised later in this section, the dialogue is ended between the agents when the signature to be mapped is empty, i.e. $\Sigma^t = \emptyset$.

Appendix D

Glossary of Terms

Chapter Outline

This Appendix (D) documents a glossary of symbols used in this thesis.

D.1 Glossary

Throughout this thesis, there have been numerous definitions and terms used to define and formalise the dialogue and corresponding decision making strategy used by the agent to achieve the goal of generating an alignment. Table D.1 presented in this appendix presents a summary of all the symbols that are used throughout this thesis.

TABLE D.1: The main symbols used throughout this thesis.

$\mathcal{T}$	Set of all legal moves τ
τ	Each movetype
Σ^t	The task signature of concepts to be mapped within the dialogue
$\mathcal{M}$	Dialogue as a sequence of moves $\mathcal{T}$
Σ_p	Agent's private ontology
Σ_d	Agent's disclosable ontology
ϖ	Triple $\langle s, p, o \rangle$
Π	Set of triples $\langle s, p, o \rangle$
$\mathcal{R}$	Ranking of triples ϖ
$\mathcal{NS}$	Set of all matching triple pairs (ϖ, ϖ')
CS	Public commitment store
Γ	Private gamma store available to each of the agents
$\mathsf{N_R}$	Relations or (*edges*) within the ontology
$\mathsf{N_C}$	Concepts or (*nodes*) within the ontology
x	Sender agent
$\hat{x}$	Receiver agent
ℓ	Formulae defined by language $\mathcal{L}$
$\mathcal{L}$	Set of formulae ℓ
Args($\mathcal{L}$)	Set of all arguments derivable from the the language $\mathcal{L}$
$\mathcal{P}$	Set of participating agents including {proponent, opponent}
σ_l	Lexical similarity metric
ϵ_l	Lexical similarity threshold
σ_n	Neighbourhood similarity metric
ϵ_n	Neighbourhood similarity threshold
σ_s	Structural similarity metric

Bibliography

[1] Rakesh Agrawal, Alexandre Evfimievski, and Ramakrishnan Srikant, *Information sharing across private databases*, Proceedings of the 2003 ACM SIGMOD international conference on Management of data, ACM, 2003, pp. 86–97.

[2] José-Luis Aguirre, Kai Eckert, Jérôme Euzenat, Alfio Ferrara, Willem Robert van Hage, Laura Hollink, Christian Meilicke, Andriy Nikolov, Dominique Ritze, François Scharffe, Pavel Shvaiko, Ondrej Sváb-Zamazal, Cássia Trojahn dos Santos, Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, and Benjamin Zapilko, *Results of the ontology alignment evaluation initiative 2012*, OM (Pavel Shvaiko, Jérôme Euzenat, Anastasios Kementsietsidis, Ming Mao, Natasha Fridman Noy, and Heiner Stuckenschmidt, eds.), CEUR Workshop Proceedings, vol. 946, CEUR-WS.org, 2012.

[3] Boanerges Aleman-Meza, C Halaschek-Weiner, I Budak Arpinar, Cartic Ramakrishnan, and Amit P Sheth, *Ranking complex relationships on the semantic web*, IEEE Internet computing **9** (2005), no. 3, 37–44.

[4] Leila Amgoud, Sihem Belabbès, and Henri Prade, *A formal general setting for dialogue protocols*, International Conference on Artificial Intelligence: Methodology, Systems, and Applications, Springer, 2006, pp. 13–23.

[5] A. Ankolekar, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, D. Martin, S. McIlraith, S. Narayanan, M. Paolucci, T. Payne, and K. Sycara, *DAML-S: Web Service Description for the Semantic Web*, First International Semantic Web Conference (ISWC) Proceedings, 2002, pp. 348–363.

[6] Amina Annane, Zohra Bellahsene, Faical Azouaou, and Clement Jonquet, *Using background knowledge to enhance ontology matching: a survey*, In submission, 2016.

[7] Michael Anslow and Michael Rovatsos, *Aligning experientially grounded ontologies using language games*, pp. 15–31, Springer International Publishing, Cham, 2015.

[8] Katie Atkinson, *What should we do?: Computational representation of persuasive argument in practical reasoning*, University of Liverpool, 2005.

[9] Katie Atkinson, Trevor Bench-Capon, and Peter Mcburney, *A dialogue game protocol for multi-agent argument over proposals for action*, Autonomous Agents and Multi-Agent Systems **11** (2005), no. 2, 153–171.

[10] John Langshaw Austin, *How to do things with words*, Oxford university press, 1975.

[11] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (eds.), *The description logic handbook: Theory, implementation, and applications*, Cambridge University Press, New York, NY, USA, 2003.

[12] Franz Baader, Ian Horrocks, and Ulrike Sattler, *Description logics as ontology languages for the semantic web*, Mechanizing Mathematical Reasoning, Springer, 2005, pp. 228–248.

[13] Sidney Bailin and Walt Truszkowski, *Ontology negotiation: How agents can really get to know each other*, Innovative Concepts for Agent-Based Systems, Springer, 2003, pp. 320–334.

[14] Sidney C Bailin and Walt Truszkowski, *Ontology negotiation: How agents can really get to know each other*, Workshop on Radical Agent Concepts, Springer, 2002, pp. 320–334.

[15] Payam Barnaghi, Amit Sheth, and Cory Henson, *From data to actionable knowledge: Big data challenges in the web of things [guest editors' introduction]*, Intelligent Systems, IEEE **28** (2013), no. 6, 6–11.

[16] Sean Bechhofer, *Owl: Web ontology language*, Encyclopedia of Database Systems, Springer, 2009, pp. 2008–2009.

[17] Alexandros Belesiotis, Michael Rovatsos, and Iyad Rahwan, *A generative dialogue system for arguing about plans in situation calculus*, International Workshop on Argumentation in Multi-Agent Systems, Springer, 2009, pp. 23–41.

[18] Paolo Besana and Dave Robertson, *Probabilistic dialogue models for dynamic ontology mapping*, pp. 41–51, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[19] Derek Bickerton, *Adam's tongue: how humans made language, how language made humans*, Hill and Wang, 2009.

[20] Christian Bizer, Tom Heath, and Tim Berners-Lee, *Linked data-the story so far*, Semantic services, interoperability and web applications: emerging concepts (2009), 205–227.

[21] Willem Nico Borst, *Construction of engineering ontologies for knowledge sharing and reuse*, Universiteit Twente, 1997.

[22] Phillip Brooker, Julie Barnett, Timothy Cribbin, and Sanjay Sharma, *Have we even solved the first 'big data challenge?' practical issues concerning data collection and visual representation for social media analytics*, pp. 34–50, Palgrave Macmillan UK, London, 2016.

[23] Anila Sahar Butt, Armin Haller, and Lexing Xie, *Dwrank: Learning concept ranking for ontology search*, Semantic Web (2016), 447–461.

[24] Michelle Cheatham and Pascal Hitzler, *String similarity metrics for ontology alignment*, International Semantic Web Conference, Springer, 2013, pp. 294–309.

[25] Paula Chocron, *Vocabulary alignment in open and heterogeneous interactions: is it possible?*, Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017) (2017), 1064–1072.

[26] Morten H Christiansen and Simon Kirby, *Language evolution*, OUP Oxford, 2003.

[27] Eve V Clark, *First language acquisition*, Cambridge University Press, 2009.

[28] Herbert H Clark and Edward F Schaefer, *Contributing to discourse*, Cognitive science **13** (1989), no. 2, 259–294.

[29] Chris Clifton, Murat Kantarcio lu, AnHai Doan, Gunther Schadow, Jaideep Vaidya, Ahmed Elmagarmid, and Dan Suciu, *Privacy-preserving data integration and sharing*, Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, ACM, 2004, pp. 19–26.

[30] Isabel F Cruz, Roberto Tamassia, and Danfeng Yao, *Privacy-preserving schema matching using mutual information*, Data and Applications Security XXI, Springer, 2007, pp. 93–94.

[31] René Descartes, *Discourse on the method: And, meditations on first philosophy*, Yale University Press, 1996.

[32] W. E. Djeddi and M. T. Khadir, *Xmap: A novel structural approach for alignment of owl-full ontologies*, 2010 International Conference on Machine and Web Intelligence, Oct 2010, pp. 368–373.

[33] Warith Eddine Djeddi and Mohamed Tarek Khadir, *A novel approach using context-based measure for matching large scale ontologies*, pp. 320–331, Springer International Publishing, 2014.

[34] William F Dowling and Jean H Gallier, *Linear-time algorithms for testing the satisfiability of propositional horn formulae*, The Journal of Logic Programming **1** (1984), no. 3, 267–284.

[35] Zlatan Dragisic, Kai Eckert, Jérôme Euzenat, Daniel Faria, Alfio Ferrara, Roger Granada, Valentina Ivanova, Ernesto Jiménez-Ruiz, Andreas Oskar Kempf, Patrick Lambrix, et al., *Results of the ontology alignment evaluation initiative 2014*, Proceedings of the 9th International Conference on Ontology Matching-Volume 1317, CEUR-WS. org, 2014, pp. 61–104.

[36] Marc Ehrig and Steffen Staab, *Qom-quick ontology mapping*, International Semantic Web Conference, Springer, 2004, pp. 683–697.

[37] Peter L Elkin, Steven H Brown, Casey S Husser, Brent A Bauer, Dietlind Wahner-Roedler, S Trent Rosenbloom, and Ted Speroff, *Evaluation of the content coverage of snomed ct: ability of snomed clinical terms to represent clinical problem lists*, Mayo Clinic Proceedings, vol. 81, Elsevier, 2006, pp. 741–748.

[38] J. Euzenat and P. Shvaiko, *Ontology matching*, Springer-Verlag, 2007.

[39] Jérôme Euzenat, *Semantic precision and recall for ontology alignment evaluation*, Proceedings of the 20th International Joint Conference on Artifical Intelligence (San Francisco, CA, USA), IJCAI'07, Morgan Kaufmann Publishers Inc., 2007, pp. 348–353.

[40] Jérôme Euzenat, *Interaction-based ontology alignment repair with expansion and relaxation*, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 185–191.

[41] Jérôme Euzenat, Pavel Shvaiko, et al., *Ontology matching*, vol. 18, Springer, 2007.

[42] Xiuyi Fan and Francesca Toni, *On computing explanations in abstract argumentation*, ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014), 2014, pp. 1005–1006.

[43] Daniel Faria, Catarina Martins, Amruta Nanavaty, Daniela Oliveira, Booma Sowkarthiga, Aynaz Taheri, Catia Pesquita, Francisco M Couto, and Isabel F Cruz, *Aml results for oaei 2015.*, OM, 2015, pp. 116–123.

[44] Daniel Faria, Catia Pesquita, Emanuel Santos, Isabel F Cruz, and Francisco M Couto, *Agreementmakerlight: a scalable automated ontology matching system*, Proceedings of the 10th International Conference on Data Integration in the Life Sciences (DILS 2014), 2014, pp. 29–32.

[45] Daniel Faria, Catia Pesquita, Emanuel Santos, Matteo Palmonari, Isabel F. Cruz, and Francisco M. Couto, *The agreementmakerlight ontology matching system*, pp. 527–541, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[46] Christiane Fellbaum, *Wordnet*, Wiley Online Library, 1998.

[47] Christiane Fellbaum, *Wordnet: An electronic lexical database*, Bradford Books, 1998.

[48] Mark D Flood, HV Jagadish, Louiqa Raschid, et al., *Big data challenges and opportunities in financial stability monitoring*, 2016.

[49] D. Gale and L. S. Shapley, *College admissions and the stability of marriage*, The American Mathematical Monthly **69** (1962), no. 1, 9–15.

[50] Pierdaniele Giaretta and N Guarino, *Ontologies and knowledge bases towards a terminological clarification*, Towards very large knowledge bases: knowledge building & knowledge sharing **25** (1995), 32.

[51] Fausto Giunchiglia, Aliaksandr Autayeu, and Juan Pane, *S-match: an open source framework for matching lightweight ontologies*, Semantic Web **3** (2012), no. 3, 307–317.

[52] Fausto Giunchiglia, Pavel Shvaiko, and Mikalai Yatskevich, *S-match: an algorithm and an implementation of semantic matching*, ESWS, vol. 3053, Springer, 2004, pp. 61–75.

[53] Fausto Giunchiglia, Pavel Shvaiko, and Mikalai Yatskevich, *Discovering missing background knowledge in ontology matching*, ECAI, vol. 141, 2006, pp. 382–386.

[54] Asuncion Gomez-Perez, Mariano Fernández-López, and Oscar Corcho, *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*, Springer Science & Business Media, 2006.

[55] Bernardo Cuenca Grau and Egor V Kostylev, *Logical foundations of privacy-preserving publishing of linked data.*, AAAI, 2016, pp. 943–949.

[56] Bernardo Cuenca Grau and Boris Motik, *Reasoning over ontologies with hidden content: The import-by-query approach*, Journal of Artificial Intelligence Research (JAIR) **45** (2012), 197–255.

[57] H Paul Grice, Peter Cole, Jerry Morgan, et al., *Logic and conversation*, 1975 (1975), 41–58.

[58] Thomas R Gruber et al., *A translation approach to portable ontology specifications*, Knowledge acquisition **5** (1993), no. 2, 199–220.

[59] Nicola Guarino et al., *Formal ontology and information systems*, Proceedings of FOIS, vol. 98, 1998, pp. 81–97.

[60] Nicola Guarino, Daniel Oberle, and Steffen Staab, *What is an ontology?*, pp. 1–17, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[61] Charles Leonard Hamblin, *Fallacies*, Methuen, 1970.

[62] Wei Hu, Ningsheng Jian, Yuzhong Qu, and Yanbing Wang, *Gmo: A graph matching for ontologies*, Proceedings of K-CAP Workshop on Integrating Ontologies, 2005, pp. 41–48.

[63] Wei Hu and Yuzhong Qu, *Falcon-ao: A practical ontology matching system*, Web Semantics: Science, Services and Agents on the World Wide Web **6** (2008), no. 3, 237–239.

[64] Joris Hulstijn, *Dialogue models for inquiry and transaction*, University of Twente, 2000.

[65] Antonio Jiménez, Mari Carmen Suárez-Figueroa, Alfonso Mateos, Asunción Gómez-Pérez, and Mariano Fernández-López, *A maut approach for reusing ontologies*, 2012 IEEE 28th International Conference on Data Engineering Workshops, April 2012, pp. 34–40.

[66] Antonio Jiménez, Mari Carmen Suárez-Figueroa, Alfonso Mateos, Asunción Gómez-Pérez, and Mariano Fernández-López, *A maut approach for reusing domain ontologies on the basis of the neon methodology*, International Journal of Information Technology & Decision Making **12** (2013), no. 05, 945–968.

[67] Ernesto Jiménez-Ruiz, *Oaei*, `http://oaei.ontologymatching.org`, 2017-6-12.

[68] Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau, *Logmap: Logic-based and scalable ontology matching*, pp. 273–288, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[69] Ernesto Jiménez-Ruiz, Terry R Payne, Alessandro Solimando, and Valentina Tamma, *Avoiding alignment-based conservativity violations through dialogue*, Proceedings OWLED 2015, 2015.

[70] Ernesto Jiménez-Ruiz, Terry R Payne, Alessandro Solimando, and Valentina Tamma, *Limiting logical violations in ontology alignment through negotiation*, Proceedings KR 2016, 2016.

[71] Charles H Kahn, *Plato and the socratic dialogue: The philosophical use of a literary form*, Cambridge University Press, 1998.

[72] Yannis Kalfoglou and Bo Hu, *Cms: Crosi mapping system-results of the 2005 ontology alignment contest*, K-Cap Integrating Ontologies (2005), 77–85.

[73] Nishan C. Karunatillake, Nicholas R. Jennings, Iyad Rahwan, and Peter McBurney, *Dialogue games that agents play within a society*, Artificial Intelligence **173** (2009), no. 9–10, 935–981.

[74] John Kelly III and Steve Hamm, *Smart machines: Ibmõs watson and the era of cognitive computing*, Columbia University Press, 2013.

[75] Anya Kim, Lance J Hoffman, and C Dianne Martin, *Building privacy into the semantic web: An ontology needed now*, Proceedings of Semantic Web Workshop, Hawaii, USA, 2002.

[76] Michel Klein, *Combining and relating ontologies: An analysis of problems and solutions*, 2001.

[77] Harold W Kuhn, *The hungarian method for the assignment problem*, Naval research logistics quarterly **2** (1955), no. 1-2, 83–97.

[78] L. Laera, I. Blacoe, V. Tamma, T.R. Payne, J. Euzenat, and T.J.M. Bench-Capon, *Argumentation over ontology correspondences in MAS*, Proceedings AAMAS 2007, 2007, pp. 1285–1292.

[79] Loredana Laera, Valentina Tamma, Jérôme Euzenat, Trevor Bench-Capon, and Terry Payne, *Arguing over ontology alignments*, Proceedings of the 1st International Conference on Ontology Matching-Volume 225, CEUR-WS. org, 2006, pp. 49–60.

[80] Loredana Laera, Valentina Tamma, Jérôme Euzenat, Trevor Bench-Capon, and Terry Payne, *Reaching agreement over ontology alignments*, The Semantic Web-ISWC 2006, Springer, 2006, pp. 371–384.

[81] Loredana Laera, Valentina Tamma, Jérôme Euzenat, Trevor Bench-Capon, and Terry Payne, *Reaching agreement over ontology alignments*, pp. 371–384, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[82] W3Schools LLC, *Ms rdfs nt*, `http://http://w3schools.sinsixx.com/rdf/rdf_schema.asp.htm`, 2017-3-30.

[83] Angela Locoro, Jérôme David, and Jérôme Euzenat, *Context-based matching: design of a flexible framework and experiment*, Journal on data semantics **3** (2014), no. 1, 25–46.

[84] Martin Loetzsch Luc Steels, *Naming game*, `http://ai.vub.ac.be`, 2017-3-30.

[85] Prodromos Malakasiotis and Ion Androutsopoulos, *Learning textual entailment using svms and string similarity measures*, Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, Association for Computational Linguistics, 2007, pp. 42–47.

[86] William C. Mann, *Dialogue games: Conventions of human interaction*, Argumentation **2** (1988), no. 4, 511–532.

[87] Peter Mcburney, Rogier M Van Eijk, Simon Parsons, and Leila Amgoud, *A dialogue game protocol for agent purchase negotiations*, Autonomous Agents and Multi-Agent Systems **7** (2003), no. 3, 235–273.

[88] John McCarthy, *Ascribing mental qualities to machines*, Philosophical Perspectives in Artificial Intelligence (Martin Ringle, ed.), Humanities Press, 1979.

[89] John McCarthy, *Artificial intelligence and philosophy*, Journal of ACM (1995), 293–320.

[90] Jarred McGinnis and David Robertson, *Realizing agent dialogues with distributed protocols.*, AC, Springer, 2004, pp. 106–119.

[91] Sergey Melnik, Hector Garcia-Molina, and Erhard Rahm, *Similarity flooding: A versatile graph matching algorithm and its application to schema matching*, Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, USA, February 26 - March 1, 2002, 2002, pp. 117–128.

[92] George A Miller, *Wordnet: a lexical database for english*, Communications of the ACM **38** (1995), no. 11, 39–41.

[93] Prasenjit Mitra, Peng Liu, and Chi-Chun Pan, *Privacy-preserving ontology matching*, AAAI Workshop on Context and Ontologies, 2005.

[94] Mark A. Musen, *Dimensions of knowledge sharing and reuse*, Computers and Biomedical Research **25** (1992), no. 5, 435–467.

[95] Mark A Musen, *National cancer institute thesaurus*, Encyclopedia of Systems Biology (2013), 1492–1492.

[96] Roberto Navigli and Simone Paolo Ponzetto, *Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network*, Artif. Intell. **193** (2012), 217–250.

[97] Robert Neches, Richard E Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, and William R Swartout, *Enabling technology for knowledge sharing*, AI magazine **12** (1991), no. 3, 36.

[98] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu, *Using of jaccard coefficient for keywords similarity*, Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, 2013.

[99] Natalya F. Noy and Deborah L. Mcguinness, *Ontology development 101: A guide to creating your first ontology*, Tech. report, Stamford, 2001.

[100] Natalya F Noy, Mark A Musen, Jose LV Mejino, and Cornelius Rosse, *Pushing the envelope: challenges in a frame-based representation of human anatomy*, Data & Knowledge Engineering **48** (2004), no. 3, 335–359.

[101] University of Manchester, *Protege tutorial*, `http://owl.cs.manchester.ac.uk/publications/talks-and-tutorials/protg-owl-tutorial`, 2017-3-30.

[102] Paul Ohm, *Changing the rules: General principles for data use and analysis*, pp. 96–111, Cambridge University Press, 2014.

[103] Simon Parsons, Peter McBurney, Elizabeth Sklar, and Michael Wooldridge, *On the relevance of utterances in formal inter-agent dialogues*, Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, ACM, 2007, p. 240.

[104] Gordon Pask, *Conversation, cognition and learning*, Amsterdam: Elsevier, 1975.

[105] Terry R. Payne and Valentina Tamma, *Negotiating over ontological correspondences with asymmetric and incomplete knowledge*, Proc. AAMAS '14, 2014, pp. 517–524.

[106] Maja Pesic and Wil MP Van der Aalst, *A declarative approach for flexible business processes management*, International conference on business process management, Springer, 2006, pp. 169–180.

[107] Roberto Poli and Johanna Seibt, *Theory and applications of ontology: Philosophical perspectives*, Springer Verlag, 2010.

[108] David Robertson, *A lightweight coordination calculus for agent systems*, International Workshop on Declarative Agent Languages and Technologies, Springer, 2004, pp. 183–197.

[109] Michael Rovatsos, *The essence project- 'towards meaning computing'*, SWeL Seminar, Heriot-Watt University, 26th October 2015.

[110] Marta Sabou, Mathieu dAquin, and Enrico Motta, *Exploring the semantic web as background knowledge for ontology matching*, Journal on data semantics XI (2008), 156–190.

[111] Emanuel Santos, Daniel Faria, Catia Pesquita, and Francisco M. Couto, *Ontology alignment repair through modularization and confidence-based heuristics*, PLOS ONE **10** (2016), no. 12, 1–19.

[112] Gabrielle Santos, Terry R. Payne, Valentina Tamma, and Floriana Grasso, *Discovering ontological correspondences through dialogue*, Knowledge Engineering and Knowledge Management: 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings 20, Springer, 2016, pp. 543–560.

[113] Gabrielle Santos, Valentina Tamma, Terry R. Payne, and Floriana Grasso, *Dialogue based meaning negotiation*, The 15th Workshop on Computational Models of Natural Argument (CMNA 2015), 2015.

[114] Gabrielle Santos, Valentina Tamma, Terry R. Payne, and Floriana Grasso, *A dialogue protocol to support meaning negotiation.: (extended abstract)*, Proceedings

of the 2016 International Conference on Autonomous Agents & Multiagent Systems (Richland, SC), AAMAS '16, International Foundation for Autonomous Agents and Multiagent Systems, 2016, pp. 1367–1368.

[115] John R Searle, *Speech acts: An essay in the philosophy of language*, vol. 626, Cambridge university press, 1969.

[116] Pavel Shvaiko and Jérôme Euzenat, *Ontology matching: state of the art and future challenges*, Knowledge and Data Engineering, IEEE Transactions on **25** (2013), no. 1, 158–176.

[117] Vassilis Spiliopoulos and George A Vouros, *Synthesizing ontology alignment methods using the max-sum algorithm*, IEEE TKDE **24** (2012), no. 5, 940–951.

[118] Luc Steels and Martin Loetzsch, *The grounded naming game*, Experiments in cultural language evolution **3** (2012), 41–59.

[119] Katherine J. Strandburg, *Monitoring, datafication, and consent: Legal approaches to privacy in the big data context*, pp. 5–43, Cambridge University Press, 2014.

[120] Rudi Studer, V. Richard Benjamins, and Dieter Fensel, *Knowledge engineering: Principles and methods*, Data Knowl. Eng. **25** (1998), no. 1-2, 161–197.

[121] Linnet Taylor, Luciano Floridi, and Bart van der Sloot, *Introduction: A new perspective on privacy*, Group Privacy, Springer, 2017, pp. 1–12.

[122] Stephen E Toulmin, *The uses of argument*, Cambridge university press, 2003.

[123] Alan M Turing, *Computing machinery and intelligence*, Mind **59** (1950), no. 236, 433–460.

[124] Michael Uschold, *Building ontologies: Towards a unified methodology*, Technical report-university of Edinburgh artificial intelligence applications institute AIAI TR **11.2** (1996), 93–136.

[125] M Ushold and R Jasper, *A framework for understanding and classifying ontology application*, Proc IJCAI99 Workshop on Ontologies and Problem-Solving Methods. Stockholm, 1999.

[126] Jurriaan van Diggelen, Robbert-Jan Beun, Frank Dignum, Rogier M. van Eijk, and John-Jules Ch. Meyer, *Anemone: an effective minimal ontology negotiation environment*, Proceedings of AAMAS 2006, 2006, pp. 899–906.

[127] Peter van Inwagen, *Being, existence, and ontological commitment*, Metametaphysics: New Essays on the Foundations of Ontology (David John Chalmers, David Manley, and Ryan Wasserman, eds.), Oxford University Press, 2009.

[128] Richard L Villars, Carl W Olofson, and Matthew Eastwood, *Big data: What it is and why you should care*, White Paper, IDC **814** (2011).

[129] Pepijn RS Visser, Dean M Jones, Trevor JM Bench-Capon, and Michael JR Shave, *Assessing heterogeneity by classifying ontology mismatches*, Proceedings of the FOIS, 1998.

[130] Pepijn RS Visser, Dean M Jones, Trevor JM Bench-Capon, and MJR Shave, *An analysis of ontology mismatches; heterogeneity versus interoperability*, AAAI 1997 Spring Symposium on Ontological Engineering, Stanford CA., USA, 1997, pp. 164–72.

[131] D. Walton and E.C.W. Krabbe, *Commitment in dialogue: Basic concepts of interpersonal reasoning*, SUNY series in Logic and Language, State University of New York Press, 1995.

[132] Eric Wanner and Lila R Gleitman, *Language acquisition: The state of the art*, CUP Archive, 1982.

[133] Paul Whitney, *The psychology of language*, Houghton Mifflin College Division, 1998.

[134] William E Winkler, *String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage.*, ERIC, 1990.

[135] Michael Wooldridge, *An introduction to multiagent systems*, John Wiley & Sons, 2009.

[136] Gang Wu, Juanzi Li, Ling Feng, and Kehong Wang, *Identifying potentially important concepts and relations in an ontology*, pp. 33–49, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[137] E. Zavitsanos, G. Paliouras, and G. A. Vouros, *Gold standard evaluation of ontology learning methods through ontology transformation and alignment*, IEEE Transactions on Knowledge and Data Engineering **23** (2011), no. 11, 1635–1648.

[138] Antoine Zimmermann, Markus Krötzsch, Jérôme Euzenat, and Pascal Hitzler, *Formalizing ontology alignment and its operations with category theory*, Proceedings of the 2006 Conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006) (Amsterdam, The Netherlands), IOS Press, 2006, pp. 277–288.